The international journal of science / 26 January 2023

nature

STOPS MAKING SENSE

Imagine a gene is a sentence starting with a capital letter and ending with a full stop, and a genome is a book telling an entire story. In some protozoan organisms, extra full stops have infiltrated the sentences, replacing specific letters of arbitrary words (namely E and W). As a r.sult, readers (in this cas. ribosomes) are confus.d as to .here thes. sentences really .nd and th. story becom.s disjointed. In this issu. of *Nature*, researchers describ. a molecular mechanism that thes. organisms must hav. developed to allo. the reader to navigat. the s.ntences as if there .ere no .xtra full stops. Th. sentences are so specifically .ncrypted that read.rs of no other organisms but thos. very fe. can decipher the story in a prop.r .ay. The trick lies in the l.ngth of the transf.r RNA molecul., and in the uniqu. modification of a singl. protein that normally ensur.s dot recognition in c.lls — the precis. end of prot.in synthesis. By the interaction of these two modified molecules, the ribosome of this protozoan knows when to correctly terminate, notwithstanding the many dots, and when to replace the dots with the two original letters (E and W), which returns meaning to the genetic information.

Engineered outcome
Supercharged immune cells gain greater range to help combat cancer

Keeping the beat
Wearable ultrasonic device offers real-time images of the heart

Get plugged in
Tools and techniques that could revamp research in 2023

VISIT…

LANZAROTE Caliente.com

Nature.2023.01.28

[Sat, 28 Jan 2023]

- **This Week**
- **News in Focus**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**

This Week

- **The water crisis is worsening. Researchers must tackle it together** [24 Jan 2023]
  • It’s unacceptable that millions living in poverty still lack access to safe water and basic sanitation. Nature Water will help researchers to find a way forward.
- **Tools such as ChatGPT threaten transparent science; here are our ground rules for their use** [24 Jan 2023]
  • As researchers dive into the brave new world of advanced AI chatbots, publishers need to acknowledge their legitimate uses and lay down clear guidelines to avoid abuse.
- **Don’t wait for an Ebola outbreak to vaccinate people against it** [24 Jan 2023]
  • The Ebola outbreak in Uganda was a wake-up call. We need preventive vaccination.

- EDITORIAL
- 24 January 2023

The water crisis is worsening. Researchers must tackle it together

It's unacceptable that millions living in poverty still lack access to safe water and basic sanitation. *Nature Water* will help researchers to find a way forward.



Communities in Gansu province in China store snow in the winter months for use in the dry summers.Credit: Hai Ying/EPA/Shutterstock

Among the world's 'poly-crises', the crisis of water is one of the most urgent. Worldwide, around 2 billion people lacked access to safe drinking water in 2020; and an estimated 1.7 billion did not have even basic

sanitation. Every year, more than 800,000 people die from diarrhoea, because of unsafe drinking water and a lack of sanitation. Most of those are in low- and lower-middle income countries. This is a mind-boggling and unacceptable situation. Even more so in an age when huge investments are being made in instant video calling, personalized medicine and talk of inhabiting other planets.

In 2015, the international community declared tackling the water crisis one of the United Nations Sustainable Development Goals (SDGs). The sixth SDG commits the world to "ensure availability and sustainable management of water and sanitation for all". But the UN acknowledges that SDG 6 is "alarmingly off track".

International diplomacy is finally starting to get its act together. In March, world leaders will assemble in New York City for the UN 2023 Water Conference. It will be the first such event in nearly half a century, a fact that by itself should shame us all.

Last October, the UN published the results of a consultation with government representatives as well as specialist and stakeholder communities on their priorities for the conference. Around 12% of respondents were from education, science and technology fields. The consensus was that data and evidence, improved access to knowledge (including Indigenous and local knowledge) and open research will be essential to getting SDG 6 back on track. Delegates attending the March conference will be looking to harness the full spectrum of established water sources and technologies, including freshwater and rainwater sources, treated groundwater, desalinated seawater and hydropower.

There's a wealth of knowledge already out there, in the form of established technologies, innovative alternatives and research that captures centuries-old knowledge and the practices of communities themselves. In the past, such knowledge has been ignored, or what has been learnt has been forgotten. Twenty years ago, for example, the UN invested in a major piece of research that captured examples of how communities living in water-stressed regions have used research and innovation to access water. The research highlighted, for example, how people in arid regions of China store snow in cellars during the winter that can then be melted for use in the summer months.

Prerequisites for tackling the water crisis include consolidating what is already known and building on that knowledge. That's why on 19 January, the Nature Portfolio of journals launched *Nature Water*. This journal will provide a space for all researchers — including those in natural and social sciences, and in engineering — to collectively contribute their knowledge, insights and the results of their learning, so that the world is on a more equitable and sustainable track. The launch issue includes research in fundamental, applied and social science, as well as opinion and analysis. Our editorial teams are committed to facilitating open science[1].

Some paths forward are clear. Damir Brdjanovic at the IHE Delft Institute for Water Education in the Netherlands writes in *Nature Water* that there's a vast body of research on alternatives to sewered sanitation — and how to use less or no water to safely dispose of faecal matter and inactivate pathogens[2]. There are alternatives to the flush toilet and underground, piped sewer networks. And Rongrong Xu at the Southern University of Science and Technology in Shenzhen, China, and colleagues report that there are ways to create hydropower, especially in Africa and Asia, without the same environmental and social impacts[3].

However, research does not exist in a vacuum. The representatives of low- and middle-income countries also want to prioritize funding. The South African government, in its response to the UN consultation, says that the annual cost to meet the SDG water and sanitation targets is between 2.3% and 2.7% of the country's gross domestic product (between US$7 billion and $7.6 billion annually). A project called the Global Commission on the Economics of Water, co-chaired by economist Mariana Mazzucato and climate scientist Johan Rockström (among others), is promising "new thinking on economics and governance" in time for the conference.

Conflict theory

Those who will be attending the conference in March also told the UN they want to see international cooperation be made a priority for water and sanitation, especially in an era of heightened geopolitical tensions. More than 25 years ago, former vice-president of the World Bank Ismail Serageldin famously wrote that twenty-first-century conflicts would be over

water, rather than oil. We are fortunate that this has not yet happened, although Serageldin told *Nature* that relations between countries that share water sources are worsening. Egypt is formally in dispute with Ethiopia over dam-building projects on the Nile River; the same is true of India and Pakistan in the Indus River Basin.

In its response to the UN, Egypt's delegation observed that the majority of people rely on water sources that are shared between nations, most of which lack formal agreements, including all-important data sharing agreements. Rhea Verbeke, at the Catholic University of Leuven in Belgium, writes in *Nature Water* of the "sobering experience" of seeing no external submissions to an open database on water purification that was created more than one year ago[4].

The delegates assembling in New York need to accept that their countries' visions will not be realized until all nations can somehow carve out a path to cooperate amid tension and conflict. Research can help to provide at least some of the right language, which is why it needs to be taken on board when decisions are being made. We in the Nature Portfolio intend to play our fullest part to make that happen.

Nature **613**, 611-612 (2023)

doi: https://doi.org/10.1038/d41586-023-00182-2

References

1. *Nature Water* **1**, 1 (2023).

2. Michalak, A. M. *et al. Nature Water* **1**, 10–18 (2023).

3. Xu, R. *et al. Nature Water* **1**, 113–122 (2023).

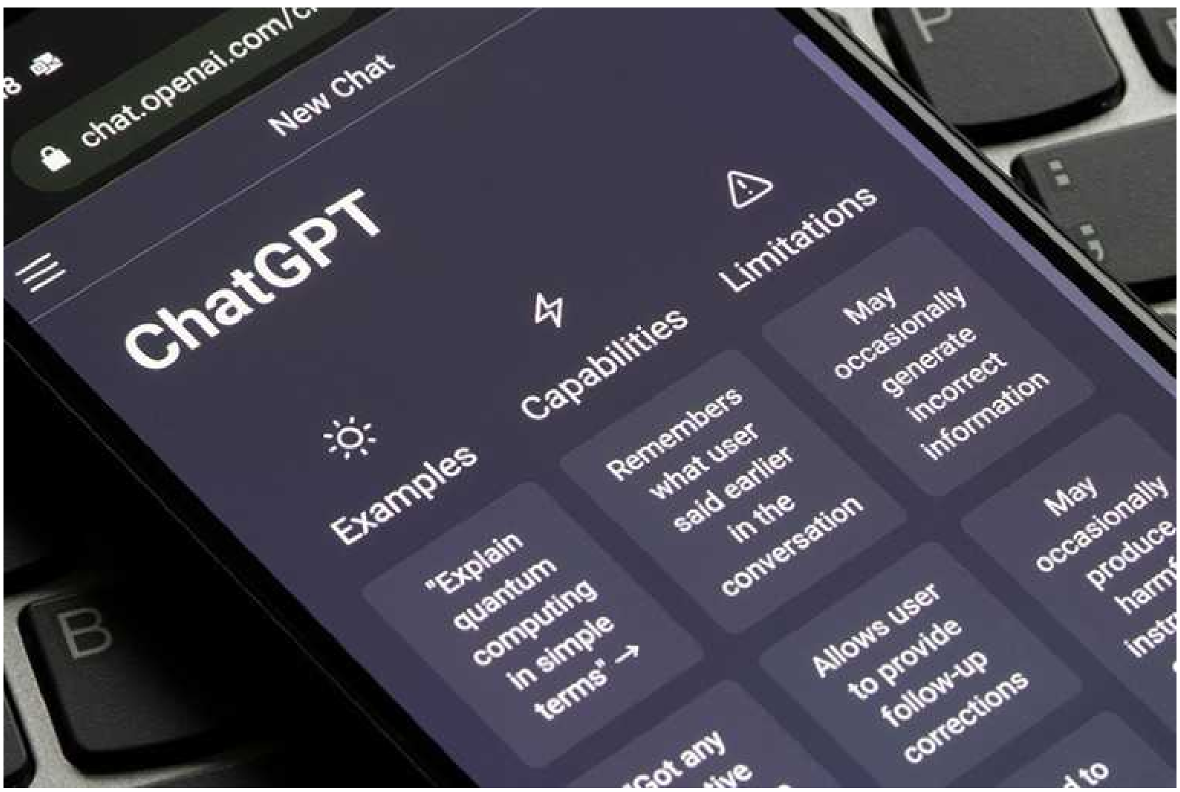
4. Verbeke, R. *Nature Water* **1**, 7–9 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00182-2

- EDITORIAL
- 24 January 2023

Tools such as ChatGPT threaten transparent science; here are our ground rules for their use

As researchers dive into the brave new world of advanced AI chatbots, publishers need to acknowledge their legitimate uses and lay down clear guidelines to avoid abuse.



ChatGPT threatens the transparency of methods that are foundational to science.Credit: Tada Images/Shutterstock

It has been clear for several years that artificial intelligence (AI) is gaining the ability to generate fluent language, churning out sentences that are increasingly hard to distinguish from text written by people. Last year, *Nature* reported that some scientists were already using chatbots as research assistants — to help organize their thinking, generate feedback on their work, assist with writing code and summarize research literature (*Nature* **611**, 192–193; 2022).

But the release of the AI chatbot ChatGPT in November has brought the capabilities of such tools, known as large language models (LLMs), to a mass audience. Its developers, OpenAI in San Francisco, California, have made the chatbot free to use and easily accessible for people who don't have technical expertise. Millions are using it, and the result has been an explosion of fun and sometimes frightening writing experiments that have turbocharged the growing excitement and consternation about these tools.



ChatGPT listed as author on research papers: many scientists disapprove

ChatGPT can write presentable student essays, summarize research papers, answer questions well enough to pass medical exams and generate helpful computer code. It has produced research abstracts good enough that scientists found it hard to spot that a computer had written them. Worryingly for society, it could also make spam, ransomware and other malicious

outputs easier to produce. Although OpenAI has tried to put guard rails on what the chatbot will do, users are already finding ways around them.

The big worry in the research community is that students and scientists could deceitfully pass off LLM-written text as their own, or use LLMs in a simplistic fashion (such as to conduct an incomplete literature review) and produce work that is unreliable. Several preprints and published articles have already credited ChatGPT with formal authorship.

That's why it is high time researchers and publishers laid down ground rules about using LLMs ethically. *Nature*, along with all Springer Nature journals, has formulated the following two principles, which have been added to our existing guide to authors (see go.nature.com/3j1jxsw). As *Nature*'s news team has reported, other scientific publishers are likely to adopt a similar stance.

First, no LLM tool will be accepted as a credited author on a research paper. That is because any attribution of authorship carries with it accountability for the work, and AI tools cannot take such responsibility.

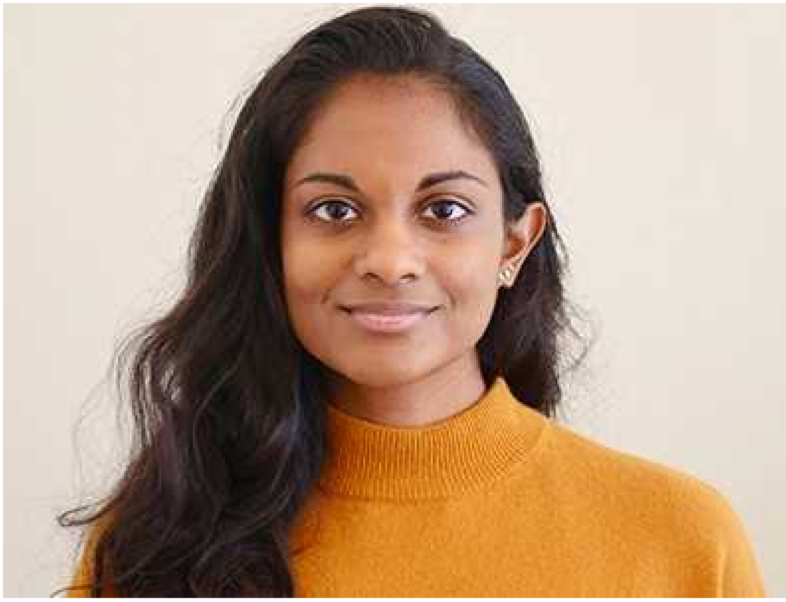
Second, researchers using LLM tools should document this use in the methods or acknowledgements sections. If a paper does not include these sections, the introduction or another appropriate section can be used to document the use of the LLM.

Pattern recognition

Can editors and publishers detect text generated by LLMs? Right now, the answer is 'perhaps'. ChatGPT's raw output is detectable on careful inspection, particularly when more than a few paragraphs are involved and the subject relates to scientific work. This is because LLMs produce patterns of words based on statistical associations in their training data and the prompts that they see, meaning that their output can appear bland and generic, or contain simple errors. Moreover, they cannot yet cite sources to document their outputs.

But in future, AI researchers might be able to get around these problems — there are already some experiments linking chatbots to source-citing tools,

for instance, and others training the chatbots on specialized scientific texts.



Don’t ask if artificial intelligence is good or fair, ask how it shifts power

Some tools promise to spot LLM-generated output, and *Nature*’s publisher, Springer Nature, is among those developing technologies to do this. But LLMs will improve, and quickly. There are hopes that creators of LLMs will be able to watermark their tools’ outputs in some way, although even this might not be technically foolproof.

From its earliest times, science has operated by being open and transparent about methods and evidence, regardless of which technology has been in vogue. Researchers should ask themselves how the transparency and trust-worthiness that the process of generating knowledge relies on can be maintained if they or their colleagues use software that works in a fundamentally opaque manner.

That is why *Nature* is setting out these principles: ultimately, research must have transparency in methods, and integrity and truth from authors. This is, after all, the foundation that science relies on to advance.

Nature **613**, 612 (2023)

doi: https://doi.org/10.1038/d41586-023-00191-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00191-1

- WORLD VIEW
- 24 January 2023

Don’t wait for an Ebola outbreak to vaccinate people against it



The Ebola outbreak in Uganda was a wake-up call. We need preventive vaccination.

- Henry Kyobe Bosa

Uganda’s most recent Ebola outbreak should be a wake-up call to the world. Last October, just 3 weeks into the outbreak, the disease spread to Kampala, a well-connected city of 1.5 million people. From there, it might easily have entered other countries.

As the incident manager at the Ugandan Ministry of Health, I coordinated all technical and operational aspects of the Ebola response. We worked tirelessly to curtail the spread, and declared the end of the outbreak on 11

January. But the quarantines and lockdowns came at a high cost, especially for the country's poorest people.

We need a better way. The world did not rely solely on lockdowns, quarantines and contact tracing to fight COVID-19. When rich countries are affected by a fatal infectious disease, there is a massive global effort to develop vaccines and therapeutics.



Vaccine incentives do not backfire — policymakers take note

Preventive, not reactive, vaccination is needed to fight Ebola. Wealthier nations must urgently fund mapping efforts and vaccine production to protect vulnerable populations in sub-Saharan Africa and, through them, the rest of the world.

Ebola outbreaks have cropped up almost every year since 2000 in Central and West Africa. Their frequency is likely to increase. With habitat destruction plus the heat, drought and other extreme weather brought on by the climate crisis, the animals that carry the virus will migrate, coming into more contact with people.

Since the first Ebola outbreaks in 1976, control has relied on the same tools: contact tracing, quarantine of confirmed cases, safe and dignified burials,

and lockdowns. These measures work best in the sparsely settled countryside, and alone are not viable in dense, complex urban settings.

I was distraught when Ebola reached Kampala. I had been in Freetown, the capital of Sierra Leone, during the West African Ebola crisis of 2014–16. Relying on contact tracing and quarantine failed us there — transmission continued for several months. I was afraid it would do so again.

Institutional quarantine, meaning separating people who have been exposed to Ebola from their families in an isolation facility for 21 days, or even just quarantining at home, can be hard enough for those who have relatives to care for, or who need those 21 days of income to feed their families. But the measures we were forced to take in Kampala included 63 days of lockdown in two rural districts with a combined population of nearly one million.



[Can mRNA vaccines transform the fight against Ebola?]

The effect on people's livelihoods and on the local timber industry was devastating. The societal disruptions that these interventions bring fuel people's anger and distrust of public-health efforts. And they are not the only way to curb the spread of the disease.

Ebola is localized in a band across Central and West Africa. High-risk regions include areas in Uganda, the Democratic Republic of the Congo,

Gabon, Sierra Leone, Liberia, South Sudan, the Ivory Coast and Ghana. Certain areas in these countries have evidence of Ebola transmission or are close to the habitats of animals that might be Ebola reservoirs, such as fruit bats. Researchers could map the populations in these countries that are at greatest risk and target them for vaccination, staving off future epidemics.

Mapping would combine three techniques: testing human blood samples for Ebola, testing blood samples from domestic and wild animals and disease modelling. For human blood samples, there is a cheap, easy way to get started: testing stored samples from nationwide HIV surveys, which many of these countries already have. For example, Uganda has the 2004–05 HIV serosurvey. We could use the results of testing these samples for Ebola to target specific geographic areas for new surveys. But African nations will need support: funding for reagents, and expertise in serosurvey design and disease modelling.

Then, we need enough Zaire-strain Ebola vaccine to vaccinate entire populations in the targeted areas. In places that have seen Ebola outbreaks, vaccine acceptance is high. Areas that have not had an outbreak will be a greater challenge, and we must invest in tailored communication and culturally sensitive ways to engage with communities to gain their trust.

A full picture of the longevity of vaccine-induced immunity is unknown for Ebola, but existing evidence is relatively encouraging (Z. A. Bornholdt and S. B. Bradfute *Lancet Infect. Dis.* **18**, 699–700; 2018). Antibody levels in the blood decrease over six months to one year, but how long T-cell immunity lasts is unknown. More data are needed, but an annual vaccine could be plausible.

Although there are approved vaccines for the Zaire strain of Ebola, this latest outbreak was caused by the Sudan strain. Comprehensive, preventive vaccination would require Sudan-strain vaccines, so global institutions need to funnel money to the entities that are developing and trialling such vaccines. More than 5,000 doses of trial vaccines arrived in Uganda just as the outbreak was ending, and the planned clinical trials must be redesigned to generate more safety and effectiveness data.

Ebola isn’t going anywhere. The wisest use of this time between outbreaks is to map high-risk areas and vaccinate those in them, instead of waiting for another outbreak. The global community has an opportunity to get it right — and to protect the world.

Nature **613**, 613 (2023)

doi: https://doi.org/10.1038/d41586-023-00184-0

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00184-0

| Section menu | Main menu |

News in Focus

- **Multimillion-dollar trade in paper authorships alarms publishers** [18 Jan 2023]
  • Journals have begun retracting publications with suspicious links to sites trading in author positions.
- **Alzheimer's drug saga prompts journal to scrutinize whistle-blowers** [12 Jan 2023]
  • Publication requires complainants to disclose financial conflicts in the wake of controversy over Cassava Sciences' experimental treatment simufilam.
- **ChatGPT listed as author on research papers: many scientists disapprove** [18 Jan 2023]
  • At least four articles credit the AI tool as a co-author, as publishers scramble to regulate its use.
- **The plan to 'Trump-proof' US science against political meddling** [12 Jan 2023]
  • Guidance document calls on agencies to draft protective scientific-integrity policies for White House review within two months.
- **China is opening up after 3 years — what does it mean for research?** [17 Jan 2023]
  • The government has abandoned its 'zero-COVID' policy and researchers from inside and outside the country are organizing to meet.
- **Researchers blast US agency's decision not to collect LGBT+ data** [13 Jan 2023]
  • Scientists call for the National Science Foundation to add a question about sexual orientation to its 2023 workforce surveys.
- **Star graveyard revealed in super-clear image of the Milky Way** [19 Jan 2023]
  • Detailed radio observations turn up previously unseen remnants of dying stars in the Galaxy.
- **The race to supercharge cancer-fighting T cells** [25 Jan 2023]

• With a slew of tools to trick out immune cells, researchers are expanding the repertoire of CAR-T therapies.

| Next section | Main menu | Previous section |

- NEWS
- 18 January 2023

Multimillion-dollar trade in paper authorships alarms publishers

Journals have begun retracting publications with suspicious links to sites trading in author positions.

- Holly Else



Buying a fraudulent authorship can cost hundreds or thousands of dollars.Credit: Getty

Research-integrity sleuths have uncovered hundreds of online advertisements that offer the chance to buy authorship on research papers to be published in reputable journals.

Publishers are investigating the claims, and have retracted dozens of articles over suspicions that people have paid to be named as authors, despite not participating in the research. Integrity specialists warn that the problem is growing, and say that other retractions are likely to follow.

Vitalij Pecharsky, a materials scientist at Iowa State University in Ames and editor-in-chief of an affected journal, was "really pissed off" when he found out about the adverts. "I know many things are for sale, but authorship? It's just crazy. This is just totally unacceptable," he told *Nature* last year (Pecharsky died in December).

Most of the adverts are posted on social-media sites including Facebook and Telegram, as well as the websites of companies that claim to offer academic publishing services. They often include the title of the paper, the journal it will be published in, the year of publication and the position of authorship slots available for purchase. Prices range from hundreds to thousands of US dollars depending on the research area and the journal's prestige.

Social-media ads

Anna Abalkina, an economist at the Free University of Berlin, and Nick Wise, an engineer at the University of Cambridge, UK, independently began investigating the adverts when they first started appearing in large numbers in 2019. Abalkina focuses on adverts on websites originating in Russia and Eastern Europe, and Wise looks at those circulated on Facebook and Telegram.

Both researchers have found published papers that they think are linked to adverts that offer authorships for sale. Two of the Telegram channels that Wise follows feature at least 300 adverts each.

In a preprint[1] posted on the arXiv server in December 2021, Abalkina describes an analysis of more than 1,000 authorship offers, together worth

more than US$6.5 million, published in 2019–21 on a Russian-language website called International Publisher. She has now linked 460 published papers to the adverts. (International Publisher did not respond to *Nature*'s request for comment.)

Spate of retractions

Journals have begun investigating and retracting papers that seem to be linked to the adverts. In July 2022, the *International Journal of Emerging Technologies in Learning* retracted 30 papers linked to adverts on International Publisher. The retraction notices say that the work was connected to a "criminal paper mill selling authorships and articles for publication in several online journals to paying customers". The blog Retraction Watch highlighted this case in an investigation into International Publisher published in December 2021.

"Generally, these things are really difficult to prove," says Tim Kersjes, a research-integrity manager at Springer Nature in Dordrecht, the Netherlands. But adverts that list specific manuscript titles offer convincing evidence that people are buying authorships, he adds. (*Nature*'s news team is independent of its publisher, Springer Nature.)

In May 2022, Springer Nature retracted a paper for the first time over suspicions that some of the authorships were paid-for, after investigating a tip-off from Abalkina. It has since retracted 11 papers in a further 5 journals over similar concerns. More investigations are under way.

"We have a zero-tolerance approach," says Kersjes. "It's a matter about integrity and reliability of the public record." Readers need to be able to trust that the authors listed on a paper actually did the work to generate the data and can be held accountable for it, he adds.

Pecharsky's publication — the *Journal of Alloys and Compounds*, published by Elsevier in Amsterdam — retracted a paper in November after a tip from Wise. Authorships on the paper had been advertised on an Iranian Telegram channel. The journal is investigating two further papers whose titles were

advertised there, and a fourth paper that shares a corresponding author with one of the papers under scrutiny.

In these cases, Pecharsky said, new authors were added to the papers after the first round of peer review. He launched an effort to review any papers published in the journal over the past two years that had authorship-change requests after the first round of review. “I just want to make sure that this thing isn’t going to propagate,” he told *Nature*.

Wise says that since October, five journals have investigated and retracted a paper each as a result of his sleuthing.

Evolving fraud

The Committee on Publication Ethics (COPE), an advisory board for publishers in Eastleigh, UK, says that authorship adverts show the evolution of paper mills — businesses that produce fake research for scientists who need publications for their CVs. “It’s big business, and run by more and more sophisticated organizations,” says Deborah Kahn, a member of the COPE board of trustees.

Abalkina thinks that the companies have developed from those that previously dealt with predatory journals — titles that publish papers for a fee but do not perform the quality-control checks of a conventional journal. Scholarly databases have removed many such journals.

The market for authorships has developed because, in many countries, researchers are still evaluated for promotion on the basis of the number of papers they publish, say Abalkina and Wise. And those seeking to buy their way onto research papers usually want them to appear in journals listed in citation databases such as Scopus and Web of Science, so that they seem legitimate.

The problem will grow, says Abalkina. “We are chasing them and they are always a step ahead.”

Nature **613**, 617-618 (2023)

doi: https://doi.org/10.1038/d41586-023-00062-9

References

1. Abalkina, A. Preprint at https://arxiv.org/abs/2112.13322 (2021).

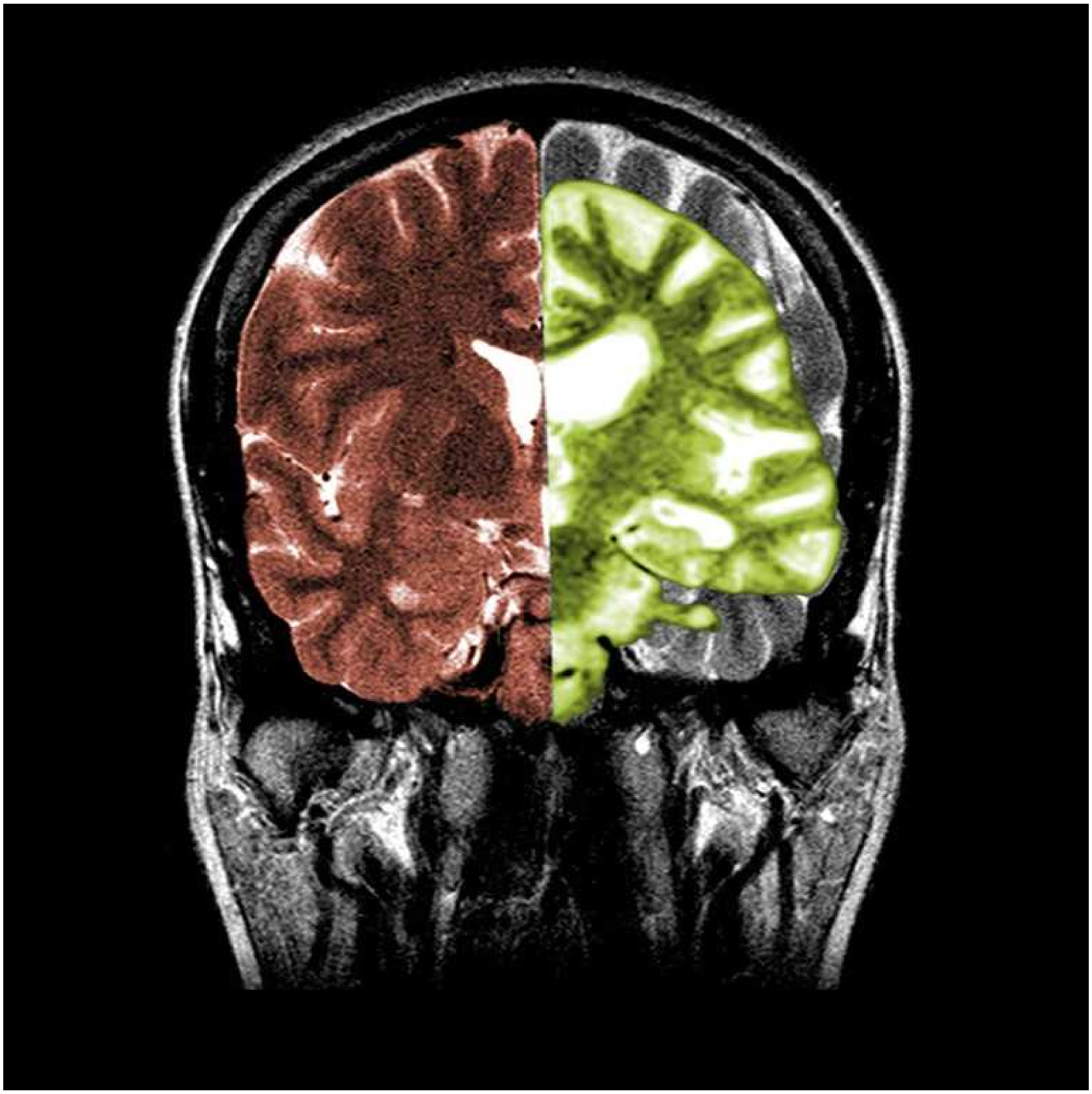
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00062-9

- NEWS
- 12 January 2023
- Correction 17 January 2023

Alzheimer’s drug saga prompts journal to scrutinize whistle-blowers

Publication requires complainants to disclose financial conflicts in the wake of controversy over Cassava Sciences’ experimental treatment simufilam.

- Holly Else



Alzheimer’s disease, as shown on the right side of this scan, transforms the brains of people with the neurodegenerative condition.Credit: Science History Images/Alamy

A scientific journal has revamped its whistle-blower policy amid a dispute over the integrity of research underlying an experimental Alzheimer’s drug.

In a 1 November 2022 editorial in *The Journal of Clinical Investigation* (*JCI*), editor-in-chief Elizabeth McNally wrote that whistle-blowers had raised concerns last year about potentially doctored images in a cache of research papers published by multiple journals, including *JCI*. Some of the

papers were related to the experimental drug simufilam, developed by biopharmaceutical firm Cassava Sciences, based in Austin, Texas, whereas others were authored by a scientist involved in the treatment's early testing.



Could drugs prevent Alzheimer's? These trials aim to find out

But, McNally says, the whistle-blowers did not disclose conflicts of interest when they lodged their complaint — and she alleges that they have profited from short-selling Cassava stock, a practice in which individuals bet that a company's stock will fall, and make money when it does. "This represents a new means of manipulating the scientific-publishing industry," she writes.

Typically, when a whistle-blower contacts a journal about concerns over manipulated images or otherwise questionable data, the allegations are taken on good faith, McNally told *Nature*. The idea that whistle-blowers could be doing this for their own financial gain "was very eye-opening to me", she says.

The group of four whistle-blowers who complained to *JCI* and other journals deny any wrongdoing and stand by their allegations, with three saying that they made only relatively small amounts of money from trading Cassava stock. When writing to the journals, they flagged possible image manipulation in 32 papers published by Hoau-Yan Wang, a medical researcher at the City University of New York (CUNY) who was responsible for much of simufilam's preclinical testing. (Five of the papers were published by Springer Nature, which also publishes *Nature*; *Nature*'s news team is independent of its publisher.)

“Our so-called conflicts of interest also have absolutely nothing to do with the objective facts in play, which are manipulated images and data,” says Adrian Heilbut, one of the whistle-blowers, who is a computational biologist and independent consultant based in New York City. Cassava denies any wrongdoing.

A drug-development controversy

The company says that simufilam, a small-molecule drug, combats Alzheimer’s by stabilizing a crucial scaffolding protein in the brain, called filamin-A, that is impaired by the disease. Other Alzheimer’s drugs use monoclonal antibodies to target clumped amyloid proteins for removal from the brains of people with the disease.

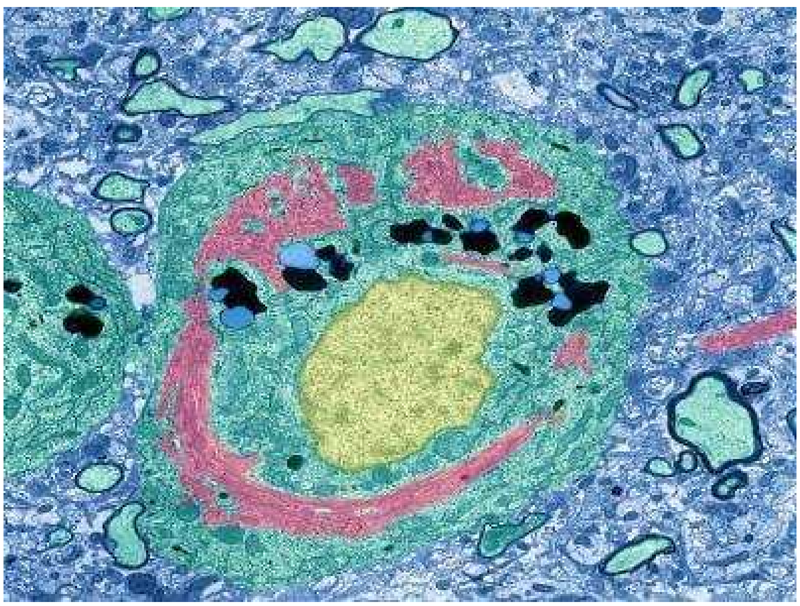


FDA approves Alzheimer’s drug lecanemab amid safety concerns

Amid the allegations about Cassava’s data, researchers have expressed concern over how simufilam works. Aside from the preliminary studies by Cassava and its collaborators, the strategy of stablilizing filamin-A to tackle Alzheimer’s hasn’t been on anyone’s radar, says George Perry, an Alzheimer’s researcher at the University of Texas at San Antonio. “The fact that it hasn’t been widely studied means that it hasn’t been confirmed.”

In a statement, a lawyer acting on behalf of Cassava told *Nature*: "The scientific community is rethinking the causes of Alzheimer's disease and looking to alternative explanations and potential solutions." Last year, two research groups, not affiliated with Cassava, published papers looking at the role that filamin-A could have in Alzheimer's disease[1,2], the statement adds.

Cassava reported interim data from a clinical trial of simufilam in a September 2021 press release. Fifty people with Alzheimer's who were treated with the small molecule for 12 months had improved cognitive test scores from when they started the trial, according to the release. The trial was 'open label', however, meaning that participants knew they were receiving the treatment, with no placebo group to ensure the improvement was caused by simufilam.



This is how an Alzheimer's gene ravages the brain

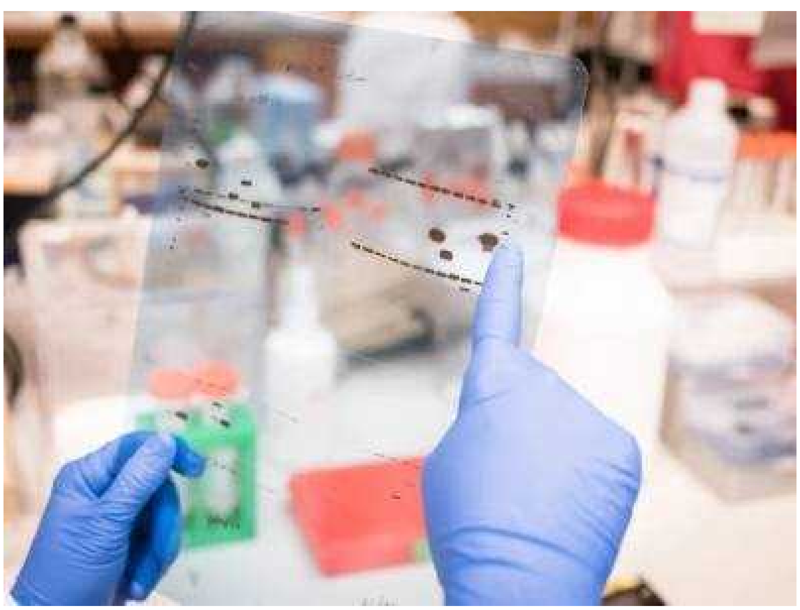
A month before the company released these results, a pair of short sellers independent of the whistle-blowers who contacted *JCI* filed a petition with the US Food and Drug Administration (FDA), pointing to potential image manipulation in Cassava's simufilam research and asking the agency to halt clinical trials. Cassava stock was at a high of US$135 per share before the petition was filed; shortly afterwards, the stock's value had plummeted by 55%.

The US Securities and Exchange Commission is investigating, according to media reports. But the FDA announced in February 2022 that it would not intervene in Cassava's clinical trials because the petition was not an appropriate way to request that the agency take this action.

The company has said that such 'citizen-petition' accusations are unsubstantiated and points out that "no government agency has accused Cassava Sciences of any type of dishonest behaviour".

Investigation ongoing

At least five papers authored by Wang and others have been retracted over concerns about image manipulation, and other investigations are ongoing. So far, journals that published two[3,4] of the three key research papers supporting simufilam as a potential Alzheimer's treatment have issued expressions of concern and corrections. These journals told *Nature* that they have not closed the cases because they are waiting for the results of a CUNY investigation into Wang's work. Cassava disputes this for one of the journals. In the case of the other, the journal says that it conducted its own investigation and found no compelling evidence of data manipulation with the intent to mislead.



Dozens of papers co-authored by Nobel laureate raise concerns

The third key paper, published in *The Journal of Prevention of Alzheimer's Disease*[5], was investigated, but one of its editors told *Nature* that it also found "no convincing evidence of manipulation of data or intent to mislead", and therefore took no action. (This journal is published by Springer Nature.)

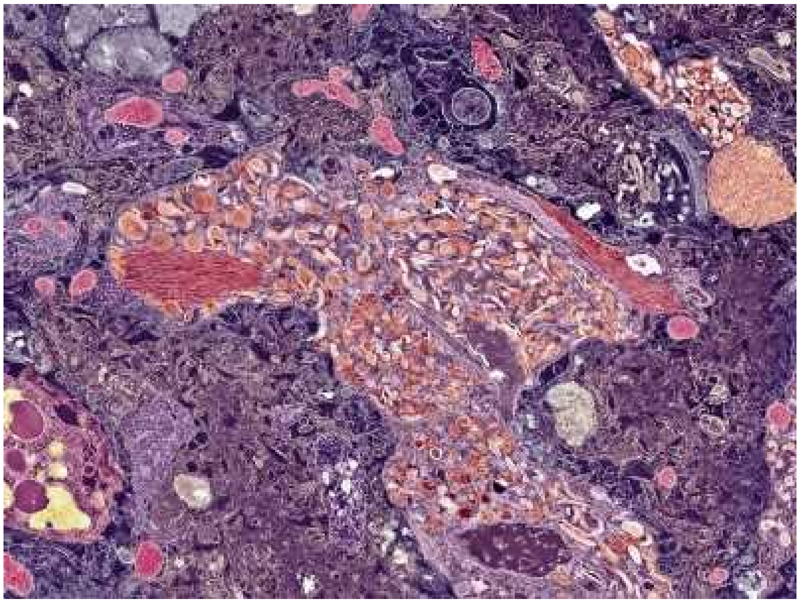
A spokesperson for CUNY told *Nature* that the university takes allegations of research misconduct seriously and cannot comment until its investigation is complete. Wang did not reply to *Nature*'s request for comment.

Last month, Cassava filed a lawsuit against the FDA petitioners, an investment company and the four whistle-blowers who contacted *JCI*: Heilbut; Jesse Brodkin, a pharmacologist who owns a preclinical-research equipment development company based in Basking Ridge, New Jersey; Enea Milioris, an independent portfolio manager who runs a life-sciences investment firm in London; and Patrick Markey, a clinical psychologist at an outpatient mental-health centre in Munich, Germany. Cassava says in its filing that the defendants spread "factually inaccurate and defamatory information".

Heilbut and Markey told *Nature* that they had no comment. Brodkin and Milioris say that instead of providing evidence to counter their claims, the company is attacking them.

Policy change

McNally says she hopes that her *JCI* editorial will remind other journals to be aware of conflicts of interest. (*JCI*'s investigation of the allegedly doctored image in its paper[6] did not substantiate the whistle-blowers' claim, she says.) "Going forward, whistleblowers, just like authors, editors, and referees, will be asked to inform us of recent, ongoing, and potential conflicts of interest," she wrote in her editorial.



[Landmark Alzheimer's drug approval confounds research community]

One journal has already said that it might follow McNally's lead. Louk Vanderschuren, editor of *Behavioural Pharmacology*, says that the journal will "definitely consider asking for disclosure of a financial conflict of interest, should an allegation of misconduct be reported to us" in future. The publication investigated allegations about one of the papers by Wang that has been scrutinized, but took no action because it received a satisfactory response from Wang, Vanderschuren told *Nature*.

Heilbut responds to McNally's action, saying he and other whistle-blowers in his group disclosed their financial positions online. Brodkin tells *Nature* that although he has no problems declaring conflicts of interest, he worries about requiring disclosures from whistle-blowers. In some cases, he says, "whistle-blowers should be granted special protections against reprisal, including an option to raise concerns anonymously".

Cassava began two phase III clinical trials in October and November last year, in which it plans to test simufilam on about 1,750 people.

Nature **613**, 618-619 (2023)

doi: https://doi.org/10.1038/d41586-023-00050-z

Updates & Corrections

- **Correction 17 January 2023**: An earlier version of this story stated that editor-in-chief Elizabeth McNally wrote in her *JCI* editorial that whistle-blowers raised concerns in August 2022. The journal received complaints starting in August 2021, but the particular instance this story is referring to occurred in July 2022.

References

1. Levert, S. *et al. Mol. Neurobiol*. https://doi.org/10.1007/s12035-022-03121-w (2022).

2. Tsujikawa, K. *et al. Sci. Adv.* **8**, eabm5029 (2022).

3. Wang, H.-Y. *et al. J. Neurobiol. Aging* **55**, 99–114 (2017).

4. Wang, H.-Y. *et al. J. Neurosci.* **32**, 9773–9784 (2012).

5. Wang, H.-Y. *et al. J. Prevention Alzheimers Dis.* **7**, 256–264 (2020).

6. Talbot, K. *et al. J. Clin. Invest.* **122**, 1316–1338 (2012).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00050-z

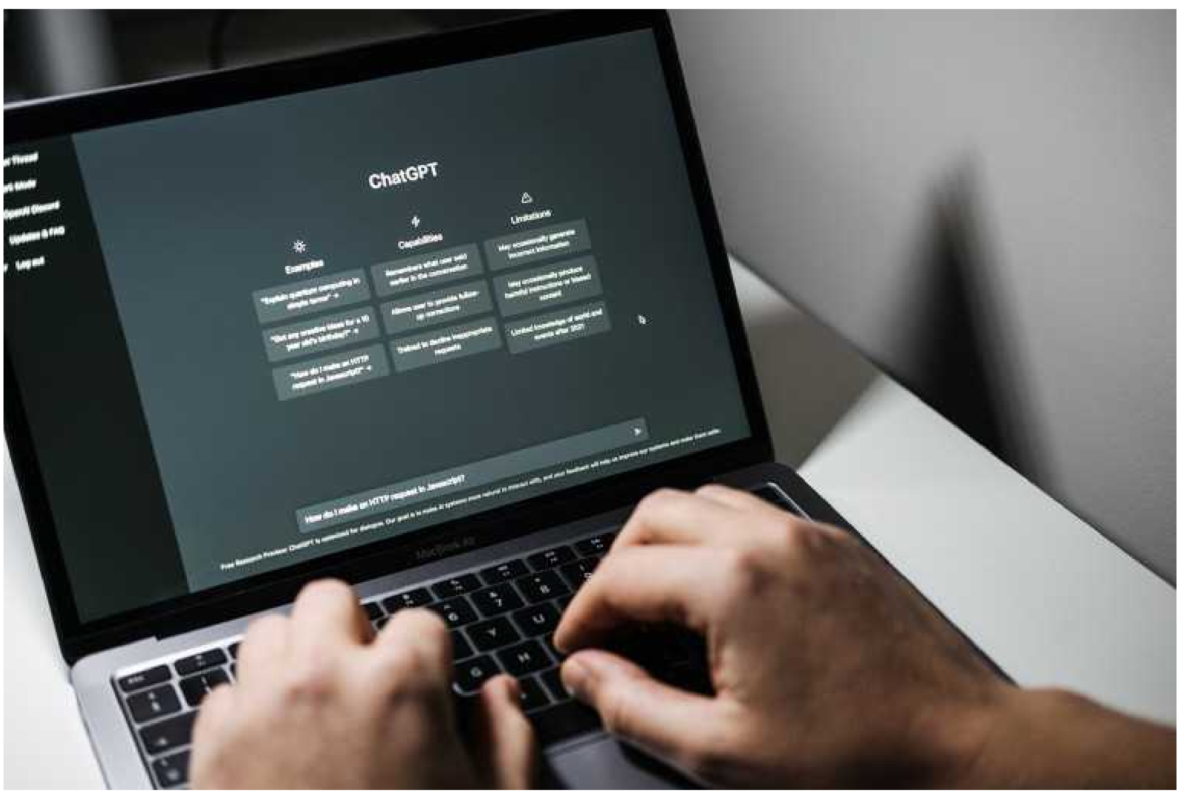
| Section menu | Main menu |

- NEWS
- 18 January 2023

ChatGPT listed as author on research papers: many scientists disapprove

At least four articles credit the AI tool as a co-author, as publishers scramble to regulate its use.

- Chris Stokel-Walker



The artificial-intelligence chatbot ChatGPT is disrupting many industries, including academia.Credit: Iryna Imago/Shutterstock

The artificial-intelligence (AI) chatbot ChatGPT that has taken the world by storm has made its formal debut in the scientific literature — racking up at least four authorship credits on published papers and preprints.

Journal editors, researchers and publishers are now debating the place of such AI tools in the published literature, and whether it's appropriate to cite the bot as an author. Publishers are racing to create policies for the chatbot, which was released as a free-to-use tool in November by tech company OpenAI in San Francisco, California.



AI bot ChatGPT writes smart essays — should professors worry?

ChatGPT is a large language model (LLM), which generates convincing sentences by mimicking the statistical patterns of language in a huge database of text collated from the Internet. The bot is already disrupting sectors including academia: in particular, it is raising questions about the future of university essays and research production.

Publishers and preprint servers contacted by *Nature*'s news team agree that AIs such as ChatGPT do not fulfil the criteria for a study author, because they cannot take responsibility for the content and integrity of scientific papers. But some publishers say that an AI's contribution to writing papers can be acknowledged in sections other than the author list. (*Nature*'s news

team is editorially independent of its journal team and its publisher, Springer Nature.)

In one case, an editor told *Nature* that ChatGPT had been cited as a co-author in error, and that the journal would correct this.

Artificial author

ChatGPT is one of 12 authors on a preprint[1] about using the tool for medical education, posted on the medical repository medRxiv in December last year.

The team behind the repository and its sister site, bioRxiv, are discussing whether it's appropriate to use and credit AI tools such as ChatGPT when writing studies, says co-founder Richard Sever, assistant director of Cold Spring Harbor Laboratory press in New York. Conventions might change, he adds.

"We need to distinguish the formal role of an author of a scholarly manuscript from the more general notion of an author as the writer of a document," says Sever. Authors take on legal responsibility for their work, so only people should be listed, he says. "Of course, people may try to sneak it in — this already happened at medRxiv — much as people have listed pets, fictional people, etc. as authors on journal articles in the past, but that's a checking issue rather than a policy issue." (Victor Tseng, the preprint's corresponding author and medical director of Ansible Health in Mountain View, California, did not respond to a request for comment.)

An editorial[2] in the journal *Nurse Education in Practice* this month credits the AI as a co-author, alongside Siobhan O'Connor, a health-technology researcher at the University of Manchester, UK. Roger Watson, the journal's editor-in-chief, says that this credit slipped through in error and will soon be corrected. "That was an oversight on my part," he says, because editorials go through a different management system from research papers.

And Alex Zhavoronkov, chief executive of Insilico Medicine, an AI-powered drug-discovery company in Hong Kong, credited ChatGPT as a co-author of a perspective article[3] in the journal *Oncoscience* last month. He

says that his company has published more than 80 papers produced by generative AI tools. “We are not new to this field,” he says. The latest paper discusses the pros and cons of taking the drug rapamycin, in the context of a philosophical argument called Pascal’s wager. ChatGPT wrote a much better article than previous generations of generative AI tools had, says Zhavoronkov.

He says that *Oncoscience* peer reviewed this paper after he asked its editor to do so. The journal did not respond to *Nature*’s request for comment.

A fourth article[4], co-written by an earlier chatbot called GPT-3 and posted on French preprint server HAL in June 2022, will soon be published in a peer-reviewed journal, says co-author Almira Osmanovic Thunström, a neurobiologist at Sahlgrenska University Hospital in Gothenburg, Sweden. She says one journal rejected the paper after review, but a second accepted it with GPT-3 as an author after she rewrote the article in response to reviewer requests.

Publisher policies

The editors-in-chief of *Nature* and *Science* told *Nature*’s news team that ChatGPT doesn’t meet the standard for authorship. “An attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs,” says Magdalena Skipper, editor-in-chief of *Nature* in London. Authors using LLMs in any way while developing a paper should document their use in the methods or acknowledgements sections, if appropriate, she says.

“We would not allow AI to be listed as an author on a paper we published, and use of AI-generated text without proper citation could be considered plagiarism,” says Holden Thorp, editor-in-chief of the Science family of journals in Washington DC.

The publisher Taylor & Francis in London is reviewing its policy, says director of publishing ethics and integrity Sabina Alam. She agrees that authors are responsible for the validity and integrity of their work, and should cite any use of LLMs in the acknowledgements section. Taylor &

Francis hasn't yet received any submissions that credit ChatGPT as a co-author.

The board of the physical-sciences preprint server arXiv has had internal discussions and is beginning to converge on an approach to the use of generative AIs, says scientific director Steinn Sigurdsson, an astronomer at Pennsylvania State University in University Park. He agrees that a software tool cannot be an author of a submission, in part because it cannot consent to terms of use and the right to distribute content. Sigurdsson isn't aware of any arXiv preprints that list ChatGPT as a co-author, and says guidance for authors is coming soon.

The ethics of generative AI

There are already clear authorship guidelines that mean ChatGPT shouldn't be credited as a co-author, says Matt Hodgkinson, a research-integrity manager at the UK Research Integrity Office in London, speaking in a personal capacity. One guideline is that a co-author needs to make a "significant scholarly contribution" to the article — which might be possible with tools such as ChatGPT, he says. But it must also have the capacity to agree to be a co-author, and to take responsibility for a study — or, at least, the part it contributed to. "It's really that second part on which the idea of giving an AI tool co-authorship really hits a roadblock," he says.

Zhavoronkov says that when he tried to get ChatGPT to write papers more technical than the perspective he published, it failed. "It does very often return the statements that are not necessarily true, and if you ask it several times the same question, it will give you different answers," he says. "So I will definitely be worried about the misuse of the system in academia, because now, people without domain expertise would be able to try and write scientific papers."

Nature **613**, 620-621 (2023)

doi: https://doi.org/10.1038/d41586-023-00107-z

References

1. Kung, T. H. *et al.* Preprint at medRxiv https://doi.org/10.1101/2022.12.19.22283643 (2022).

2. O'Connor, S. & ChatGPT *Nurse Educ. Pract.* **66**, 103537 (2023).

3. ChatGPT & Zhavoronkov, A. *Oncoscience* **9**, 82–84 (2022).

4. GPT, Osmanovic Thunström, A. & Steingrimsson, S. Preprint at HAL https://hal.science/hal-03701250 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00107-z

- NEWS
- 12 January 2023

The plan to ‘Trump-proof’ US science against political meddling

Guidance document calls on agencies to draft protective scientific-integrity policies for White House review within two months.

- Jeff Tollefson



President Joe Biden promised to restore scientific integrity to the US government when he began his tenure in the White House.Credit: Getty

The administration of US President Joe Biden has unveiled its plan to protect government science from political interference. Guidance released by the White House on 12 January lays out the standards for policies that federal agencies have been asked to develop in the coming months.

“Upholding the highest standards of scientific integrity across the Federal government is of vital importance for keeping all of America safe, healthy, and secure,” Arati Prabhakar, Biden’s science adviser and director of the White House Office of Science and Technology Policy (OSTP), wrote in an accompanying memorandum to federal agencies. “Scientists must have a seat at the table and scientific information must reach decisionmakers without inappropriate influence.”



Researchers frustrated as Biden’s science adviser resigns amid scandal

Arriving at the beginning of Biden’s third year in the White House, the document was crafted partly in response to the downplaying of science and sidelining of scientists at multiple federal agencies during the administration of former president Donald Trump. It also arrives in the wake of controversy in Biden’s own administration. Eric Lander, the president’s former science adviser and OSTP director, stepped down in February 2022 following media reports that he mistreated staff. And in August last year, the US National Academy of Sciences penalized Jane Lubchenco, the OSTP’s deputy director for climate and the environment, for violating conflict-of-interest rules by editing a paper in the journal *Proceedings of the National Academy of Sciences* that was co-authored by a former student, who is now her brother-in-law.

The OSTP did not immediately respond to requests for comment. Lubchenco has apologized for the “error in judgment”, which occurred before she joined the White House.

Government watchdogs praised the White House’s 66-page guidance document as a major — and long-overdue — step forward. However, they say, further steps will be needed to secure the role of scientists and science in government decision-making and to prevent the type of political meddling that occurred under Trump.

“This should really be heralded as a great achievement for federal science, but there is always room for improvement,” says Jacob Carter, research director for the Center for Science and Democracy at the Union of Concerned Scientists, an advocacy group based in Cambridge, Massachusetts.

Restoring trust

Biden began his tenure in the White House with a promise to restore scientific integrity and to promote science-based policymaking across the federal government. He wanted to prevent the kind of meddling that had occurred under Trump, including altering public-health recommendations from the US Centers for Disease Control and Prevention during the COVID-19 pandemic, and an incident in which Trump made false statements about a 2020 hurricane forecast that were backed up by officials he had appointed to the US National Oceanic and Atmospheric Administration.



After Trump, US researchers urge Biden to block political meddling in science

Following an initial report issued in January 2022, on how to protect scientific integrity in the government, the OSTP has now released a framework that outlines ways to strengthen, expand and, to a degree, standardize scientific-integrity policies across agencies.

It's not the first time that the White House has taken action on scientific integrity: the administration of former president Barack Obama made a similar push in 2010, and 24 federal agencies crafted their own policies at that time. However, the policies varied greatly in terms of quality and detail, and some observers have said the Obama administration didn't do enough to bolster and guide implementation after that initial push.

"It's only now in this administration that we are really seeing the follow-through that this issue needs," says Francesca Grifo, who is in charge of scientific-integrity policy at the US Environmental Protection Agency. "This is really new territory."



Four ways Trump has meddled in pandemic science — and why it matters

John Holdren, who led the effort under Obama as OSTP director, said in an e-mail to *Nature* that the agency worked hard around 2010 to ensure that strong scientific-integrity policies were crafted and properly implemented — efforts that subsequently stalled under the Trump administration. “It’s great that the Biden administration has restored the proper focus on scientific integrity that good government requires,” he added.

The new guidance includes a model scientific-integrity policy for federal agencies to use as a template. Agencies now have 60 days to submit their own policies for review by the OSTP, and are expected to post their proposed policies for public comment within 6 months. Agencies then have another three months to finalize their policies. The plan also calls for the National Science and Technology Council to establish a panel, composed of scientific-integrity officials from various agencies, as well as White House officials, which would have the authority to review agencies’ policies. The panel would also investigate potential violations by senior officials and political appointees.

Details to come

There is a lot to like about the OSTP's new guidance, advocates say, but the framework is vague about things such as the enforcement of sanctions for people who break the rules, as well as the protocol for scientists seeking to talk to journalists or the public about their work. Those details will all be left to each agency's discretion.



[How to protect US science from political meddling after Trump]

"I think that's a missed opportunity," says Lauren Kurtz, executive director of the Climate Science Legal Defense Fund, an advocacy group based in New York City. Although the policy seeks to strengthen scientific-integrity operations and create an independent panel that could provide consistency from administration to administration, Kurtz warns that these efforts aren't guaranteed to withstand future political meddling. A hostile administration could wipe out the entire framework — as well as individual agencies' policies — with the stroke of a pen.

A permanent solution would be to enact a federal law that would make it illegal for government officials now and in the future to violate scientific-integrity policies. Thus far, efforts to introduce such a bill in Congress have fallen short, but Carter remains hopeful.

"Eventually, all of this has to be codified in legislation," he says.

Nature **613**, 621-622 (2023)

doi: https://doi.org/10.1038/d41586-022-03307-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-03307-1

- NEWS
- 17 January 2023

China is opening up after 3 years — what does it mean for research?

The government has abandoned its ‘zero-COVID’ policy and researchers from inside and outside the country are organizing to meet.

- Smriti Mallapaty



Travel to and from China is increasing after quarantines for travellers were lifted this month.Credit: Andy Wong/AP/Shutterstock

Palaeoanthropologist Clément Zanolli has been waiting more than three years to see ancient teeth, unearthed in China, that could belong to a previously undescribed ape. Zanolli, from the University of Bordeaux, France, had to cancel his travel plans when China shut its borders in response to the COVID-19 pandemic in early 2020. But with the government lifting its onerous quarantine requirements for incoming travellers this month, Zanolli is one of many researchers planning their return to China for conferences, face-to-face meetings and fieldwork.

It's "fantastic that the borders are finally open", says Alice Hughes, a conservation biologist at the University of Hong Kong, who hopes to travel to the Chinese mainland around April. This is long overdue, she says.

China's strict 'zero-COVID' policy to quash the spread of the disease was among the longest and harshest worldwide. For almost three years, citizens endured mandatory testing, whole-city lockdowns and lengthy quarantines. The restrictions meant that researchers in China were largely confined to their campuses because regional and international travel was time-consuming, unpredictable and often expensive. Meanwhile, most researchers abroad were locked out of China.

But the policy's end in December has led to a surge in infections across the country. This has left some researchers cautious about travelling to China. Although they hope outbreaks will slow down soon.

Meet and greet

Many researchers in China are planning to attend conferences abroad later this year now that quarantines for returning travellers have been lifted. Jingbo Cui, an economist who studies environmental issues at Duke Kunshan University, China, wants to attend meetings in the United States and the United Kingdom to help rebuild his international reputation. "Our research needs exposure." And geneticist Shuhua Xu, at Fudan University in Shanghai, hopes to use a grant he received before the pandemic for an academic exchange with researchers at the University of Cambridge, UK.

Researchers outside China are also making plans to visit their Chinese colleagues. Virologist Linfa Wang’s top priority is to meet with lab groups in Beijing and Guangzhou that will focus on preparing for pandemics. “We can now plan with a bit more certainty,” says Wang, at Duke–National University of Singapore Medical School.

“Many people are eager to re-engage with China and the scientists here,” says Thomas Stidham, a palaeontologist at the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, in Beijing. Researchers in China have made big fossil discoveries over the past three years and palaeontologists are probably eager to get a first-hand look, says Stidham.

International collaborations dwindled during the pandemic, but a return to in-person meetings should help revive them, says Caroline Wagner, a science-policy researcher at the Ohio State University in Columbus.

Flights into and out of China are expected to increase from their pandemic levels, which will boost research connections, says Li Tang, a science- and innovation-policy researcher at Fudan University. However, flights are expensive at the moment and limited flight connections might make people reluctant to travel to and from China in the immediate future.

Persistent barriers

For some researchers, the excitement of China opening up is tempered by geopolitical tensions. The United States, Europe, the United Kingdom and Australia have increased their focus on national security and competitiveness, which might deter researchers in those regions from collaborating with their Chinese peers, says James Wilsdon, a science-policy researcher at University College London.

And in recent weeks, some countries, such as South Korea, the United States and Australia, have introduced testing requirements and other restrictions on travellers entering their borders from China because of the huge wave of COVID-19 infections there. China, in turn, has said it will not be issuing visas for travellers from some of those countries.

The pandemic has also politicized specific areas of research, including the origins of SARS-CoV-2, making scientific engagement difficult. Edward Holmes, a virologist at the University of Sydney in Australia, who publically shared[1] the SARS-CoV-2 genome in January 2020 with his long-time Chinese collaborator, virologist Zhang Yongzhen, has no plans to visit China any time soon because he fears his work there could be censored. “That makes me feel very uneasy,” says Holmes.

“Chinese society has been under COVID isolations for much longer, with much tighter censorship,” says Joy Zhang, a sociologist at the University of Kent in Canterbury, UK. “I suspect I will have a small ‘culture shock’ when I visit China next time.”

Still, researchers in China are relieved that domestic travel will now be easier. Aaron Irving, an infectious-diseases researcher at Zhejiang University in Haining, China, hopes to visit caves to trap bats coming out of hibernation to study, among other things, the bacteria and viruses they harbour, after having to cancel two trips owing to COVID-19 outbreaks. “All of us were glad that the restrictions were lifted because it makes day-to-day life much easier.”

Nature **613**, 622 (2023)

doi: https://doi.org/10.1038/d41586-023-00091-4

References

1. Zhang, Y. Z. & Holmes, E. C. *Cell* **181**, 223-227 (2020).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00091-4

- NEWS
- 13 January 2023

Researchers blast US agency's decision not to collect LGBT+ data

Scientists call for the National Science Foundation to add a question about sexual orientation to its 2023 workforce surveys.

- Max Kozlov



Lockers in a rainbow of colours at the US National Science Foundation Ice Core Facility in Lakewood, Colorado.Credit: Jim West/Alamy

The US National Science Foundation (NSF) has decided not to include a question about sexual orientation on one of its workforce surveys, setting off a social-media firestorm. More than 1,700 researchers have now signed an open letter urging the agency's director to reconsider the decision. They argue that it is crucial to collect such information to understand the make-up of the scientific community and craft policies that lessen disparities for people from sexual and gender minorities.



How LGBT+ scientists would like to be included and welcomed in STEM workplaces

The National Center for Science and Engineering Statistics (NCSES), a subdivision of the NSF, regularly administers several surveys that provide key data about the US scientific workforce to policymakers and researchers on demographics such as sex, race and ethnicity, scientific discipline and citizenship status. The NCSES has for years been considering adding questions about sexual orientation and gender identity to its surveys, but delays have frustrated scientists.

In an application filed with the US Office of Management and Budget (OMB) on 14 December 2022, the NSF said that it would include a two-part question about gender identity, but not one about sexual orientation, on its 2023 National Survey of College Graduates. Julia Milton, a spokesperson for NCSES in Alexandria, Virginia, said in a statement that data-quality and

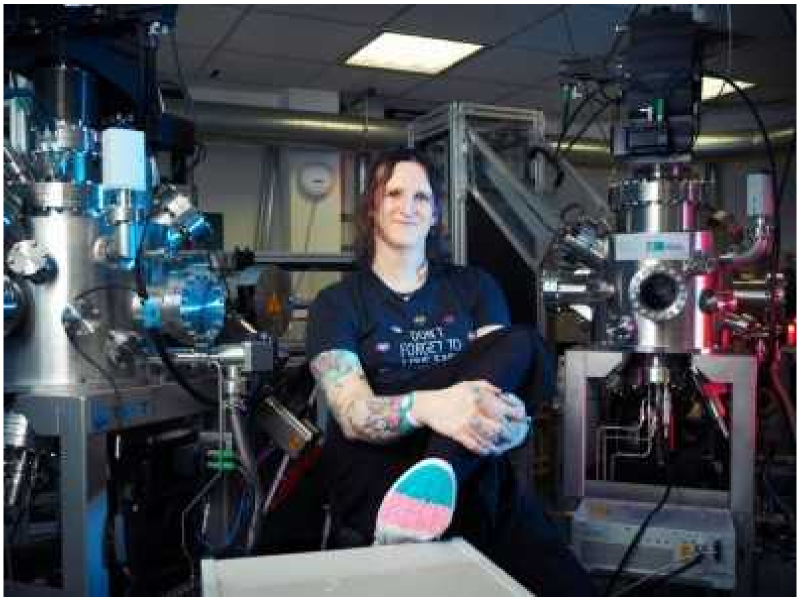
privacy issues led the agency to exclude the sexual-orientation question. The agency tested the questions during a pilot survey, and concluded that the sexual-orientation one led to more respondents quitting and not responding to the query than did other queries.

Jon Freeman, a neuroscientist at Columbia University in New York City who authored the open letter, has been pushing the NSF to include such questions for years. He is not convinced by the agency's rationale to exclude the sexual-orientation question. The NSF "justified this move with flawed analyses, inappropriate benchmarks and selective reporting of its own pilot data", Freeman writes in the letter. Milton did not answer *Nature*'s enquiries about the allegations in the letter.

"It's perplexing they would leave the sexual-orientation question off, given that it's been asked successfully for ten years now" on surveys by other federal agencies, says Nancy Bates, a former US Census Bureau statistician who co-chaired a 2022 National Academies of Science, Engineering and Medicine consensus report on measuring sex, gender identity and sexual orientation.

The 'gold standard'

The decision comes as some studies have suggested that people who identify as lesbian, gay, bisexual, transgender or another sexual or gender minority (LGBT+) are under-represented in the sciences and face more barriers and workplace harassment than those who do not identify as such[1,2].



‘I was no longer pretending or hiding’: a trans scientist finds a lab to call home

The NSF’s surveys are considered to be the “gold standard of data collection” on the scientific workforce, so this decision is highly consequential, Bates says. Although Milton says that NCSES’s surveys do not have “bearing on the determination of eligibility for federal research funding”, Emma Wojtowicz, a spokesperson for the US National Institutes of Health (NIH), notes that some agencies such as the NIH use data from the NSF’s workforce surveys to determine which groups are under-represented and therefore qualify for grants that promote diversity.

“If we don’t have the data to understand potential differences or disadvantages, the silence of the data itself reproduces inequality,” says Erin Cech, a sociologist at the University of Michigan in Ann Arbor who studies LGBT+ representation in science and who has signed the letter. In June 2022, US President Joe Biden issued an executive order directing federal agencies to “advance the responsible and effective collection” of gender-identity and sexual-orientation data.

Survey design concerns

To ensure that adding or revising questions on its surveys wouldn't result in people quitting them or otherwise distorting the data collected, NCSES ran a pilot of its college-graduates survey in 2021 with 5,000 respondents. The survey included both a gender-identity and a sexual-orientation question. In its submission to the OMB, the NCSES said that it had concerns about how the sexual-orientation question performed.

To fix LGBTQ+ disparities in science, we need the data

But the agency tested a strange version of the sexual-orientation question, says Kerith Conron, research director at the Williams Institute in Los Angeles, California, a think tank that conducts research on sexual-orientation and gender-identity law and public policy. The question began with a phrase that Bates says might have confused people, and included more response options than are used in other federal surveys. "I've been around a long time, and I've never seen that question before," Conron says. Milton replies that the expanded response options could "inform the broader federal government's effort to measure and understand sexual minorities".

Conron applauds the agency's inclusion of a gender-identity question, but says that it was a "fail on the part of NSF not to include more traditional measures of sexual orientation". In the open letter, Freeman calls on the agency to release the full data of the pilot study.

Whether or not the NSF will exclude the sexual-orientation question from its other scientific-workforce surveys is still unclear. Freeman says it would be unusual for the agency to exclude it from the college-graduates survey and then add it to other 2023 surveys.

Members of the public have until 14 January to submit comments about the NSF's data-collection plans for the survey. The OMB will then have about 30 days to review the NCSES's materials and all comments, after which it can either approve the plan or ask the NSF to revise it.

Nature **613**, 624 (2023)

doi: https://doi.org/10.1038/d41586-023-00082-5

References

1. Cech, E. A. & Waizunas, T. J. *Sci. Adv.* **7**, eabe0933 (2021).
2. Cech, E. A. & Pham, M. V. *Soc. Sci.* **6**, 12 (2017).

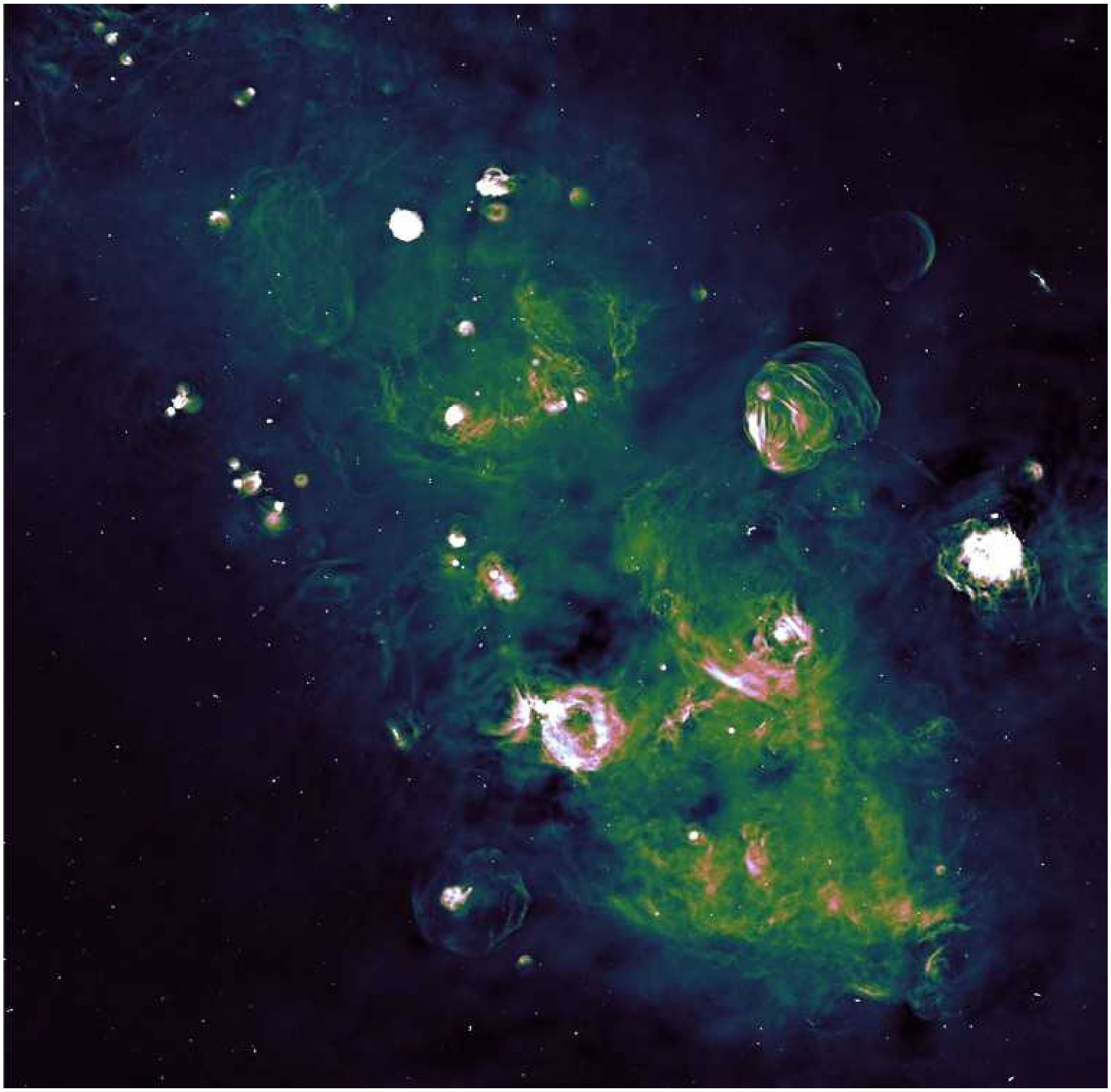
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00082-5

- NEWS
- 19 January 2023
- Correction 20 January 2023

Star graveyard revealed in super-clear image of the Milky Way

Detailed radio observations turn up previously unseen remnants of dying stars in the Galaxy.

- Jonathan O'Callaghan



The image shows five previously hidden supernova remnants. The colours represent heat, with purple denoting the coolest regions, followed by blue, green and red, with white highlighting the hottest parts. Credit: R. Kothes (NRC) and the PEGASUS team

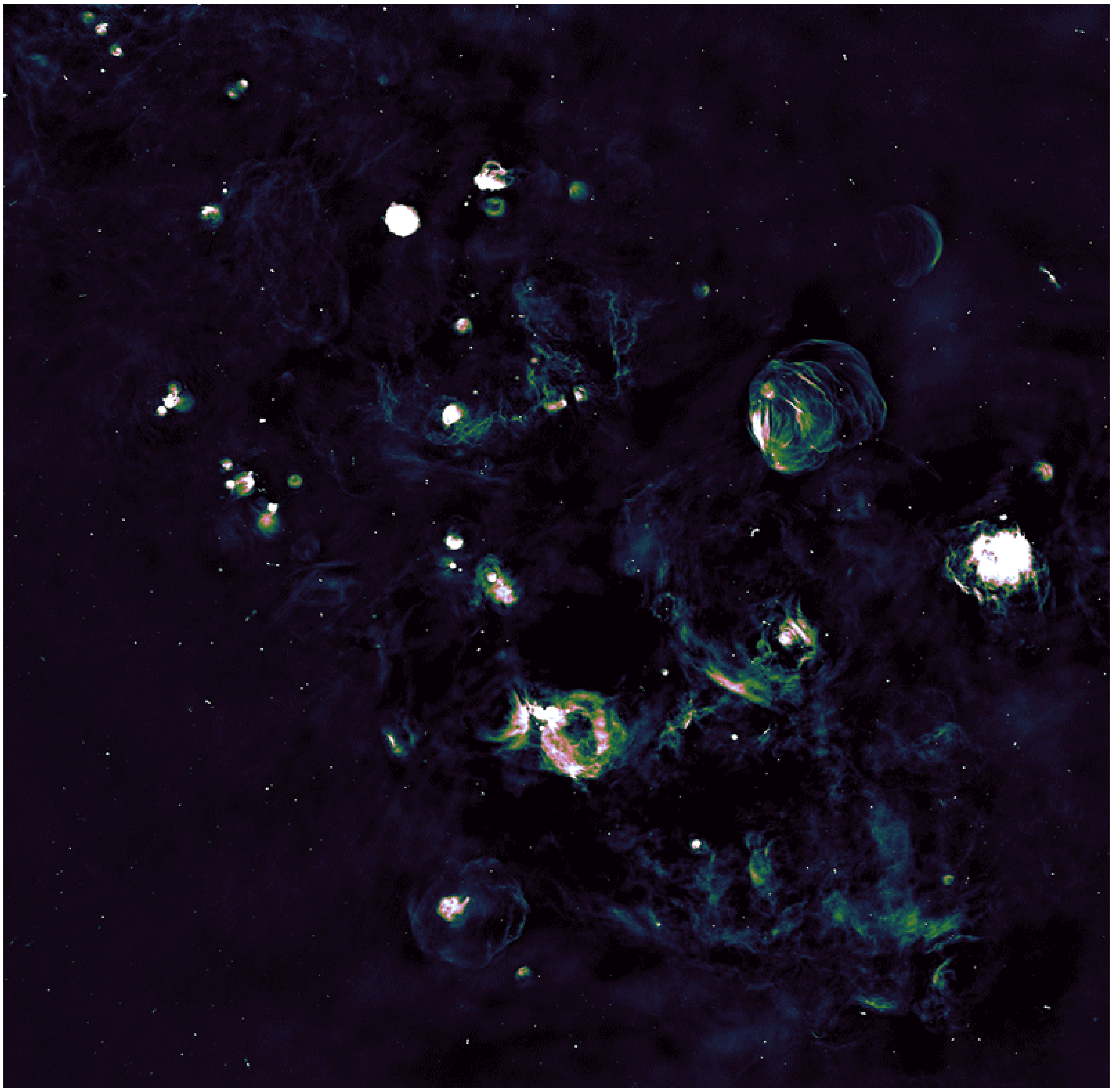
Astronomers have discovered the remains of nearly two dozen exploding stars in the Milky Way, thanks to detailed radio observations that could unveil many more such events in the Galaxy.

A star in the Milky Way is expected to explode as a supernova roughly at least once every 100 years. These violent explosions — the dramatic final throes of massive stars as they exhaust their fuel — can eject vast clouds of

dust and gas to locations many light years from the star. Such ‘supernova remnants’ can persist for thousands of years before dissipating. Studying these remnants can reveal useful information about the Galaxy, because they often contain heavy elements that give rise to other stars, planets and even life itself.

Hundreds of such remnants have been found across the Milky Way, but astronomers think that they have observed only about one-fifth of the total number. Most are found by detecting radio emissions from the remnants as they expand, revealing their otherwise invisible shapes, but many are too faint to be picked up. “There’s this missing supernova-remnant problem,” says Brianna Ball, who studies astronomy at the University of Alberta in Edmonton, Canada. “We know how many we should see, and we see a lot fewer than that.”

On 16 January, however, a project led by Ball revealed a new way to track supernova remnants down. It combined the observing power of the Australian Square Kilometre Array Pathfinder (ASKAP), a radio telescope comprising 36 antennas in Western Australia, with that of the Parkes Observatory, a single dish in New South Wales, Australia, to detect previously unseen supernova remnants in a patch of night sky.



The images from the Parkes and ASKAP radio telescopes were combined to provide detailed views of the supernova remnants.Credit: CSIRO, R. Kothes (NRC), E. Carretti (INAF) and the EMU and POSSUM teams

“We discovered 21 new candidates,” says Roland Kothes, a radio astronomer at the National Research Council of Canada, who is based in Penticton and Ball’s supervisor. “This image was the first test that we did, and it worked spectacularly well,” he says. The work has yet to be published.

The image the team released contains about five of the 21 newly discovered supernova remnants, including one shaped like a figure of eight, seen faintly

in the top left. The picture was taken along one of the spiral arms of the Milky Way, the Norma Arm, near the dense Galactic Centre, where dust and gas heavily obscure visible light.

Finding new remnants can tell us more about “the kinds of stars that explode as supernovae”, says Carlos Badenes, an astronomer at the University of Pittsburgh in Pennsylvania, who studies supernova remnants. “It’s really fantastic that these guys are doing this.”

ASKAP began a five-year survey of the entire sky of the Southern Hemisphere, which accounts for half of the visible Galaxy, last year. That should turn up many more supernova remnants. “We’re detecting sources that previous telescopes might not have been able to detect because they didn’t have the resolution or sensitivity,” says Ball. “We’re hoping it will uncover a large population.”

Nature **613**, 625 (2023)

doi: https://doi.org/10.1038/d41586-023-00110-4

Updates & Corrections

- **Correction 20 January 2023**: An earlier version of this story misstated the location of Roland Kothes.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00110-4

- NEWS FEATURE
- 25 January 2023
- Correction 27 January 2023

The race to supercharge cancer-fighting T cells

With a slew of tools to trick out immune cells, researchers are expanding the repertoire of CAR-T therapies.

- Heidi Ledford

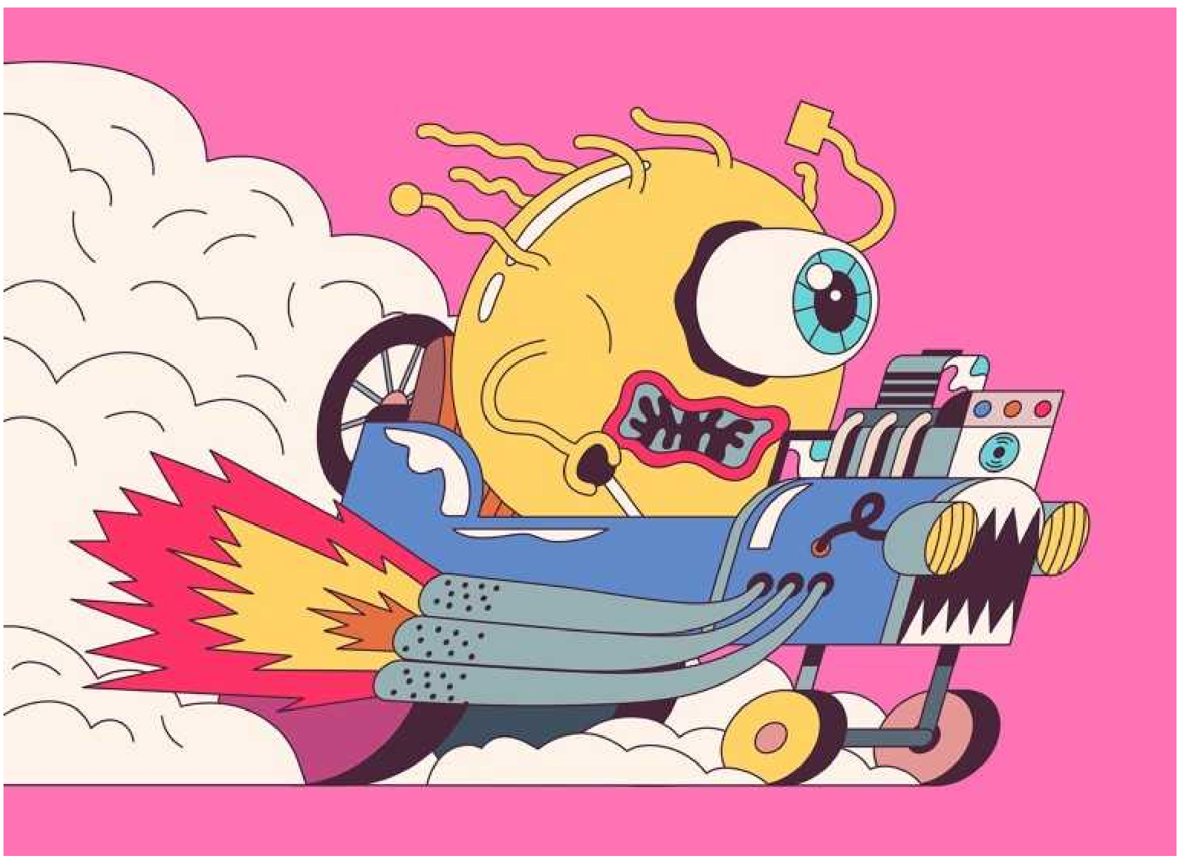
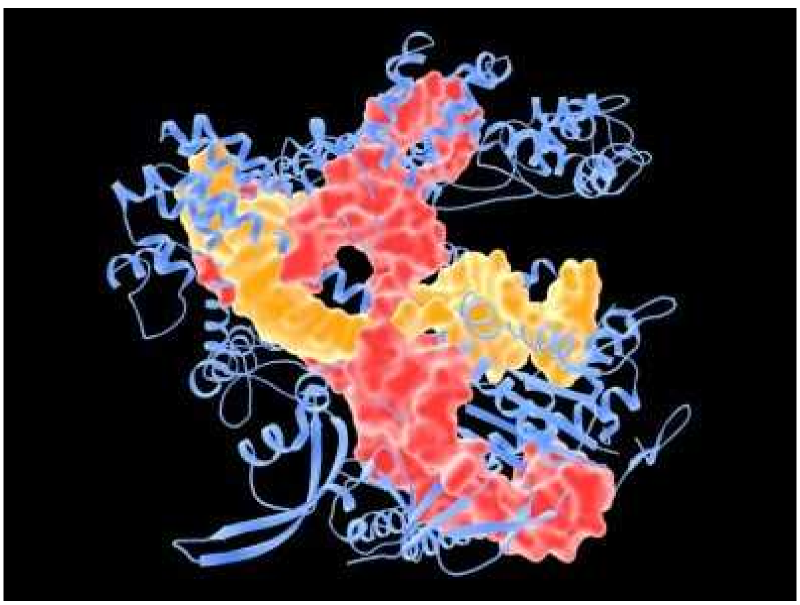


Illustration by Fabio Buonocore

Crystal Mackall remembers her scepticism the first time she heard a talk about a way to engineer T cells to recognize and kill cancer. Sitting in the audience at a 1996 meeting in Germany, the paediatric oncologist turned to the person next to her and said: “No way. That’s too crazy.”

Today, things are different. “I’ve been humbled,” says Mackall, who now works at Stanford University in California developing such cells to treat brain tumours. The US Food and Drug Administration approved the first modified T cells, called chimeric antigen receptor (CAR)-T cells, to treat a form of leukaemia in 2017. The treatments have become game changers for several cancers. Five similar products have been approved, and more than 20,000 people have received them. A field once driven by a handful of dogged researchers now boasts hundreds of laboratory groups in academia and industry. More than 500 clinical trials are under way, and other approaches are gearing up to jump from lab to clinic as researchers race to refine T-cell designs and extend their capabilities. “This field is going to go way beyond cancer in the years to come,” Mackall predicts.



CRISPR cancer trial success paves the way for personalized treatments

Advances in genome editing through processes such as CRISPR, and the ability to rewire cells through synthetic biology, have led to increasingly elaborate approaches for modifying and supercharging T cells for therapy. Such techniques are providing tools to counter some of the limitations of

current CAR-T therapies, which are expensive to make, can have dangerous side effects, and have so far been successful only against blood cancers. “These techniques have expanded what we’re able to do with CAR strategies,” says Avery Posey, a cancer immunology researcher at the University of Pennsylvania in Philadelphia. “It will really take this type of technology forward.”

Even so, the challenge of making such a ‘living drug’ from a person’s cells extends beyond complicated designs. Safety and manufacturing problems remain to be addressed for many of the newest candidates. “There’s an explosion of very fancy things, and I think that’s great,” says immunologist Michel Sadelain at the Memorial Sloan Kettering Cancer Center in New York City. “But the complexity cannot always be brought as described into a clinical setting.”

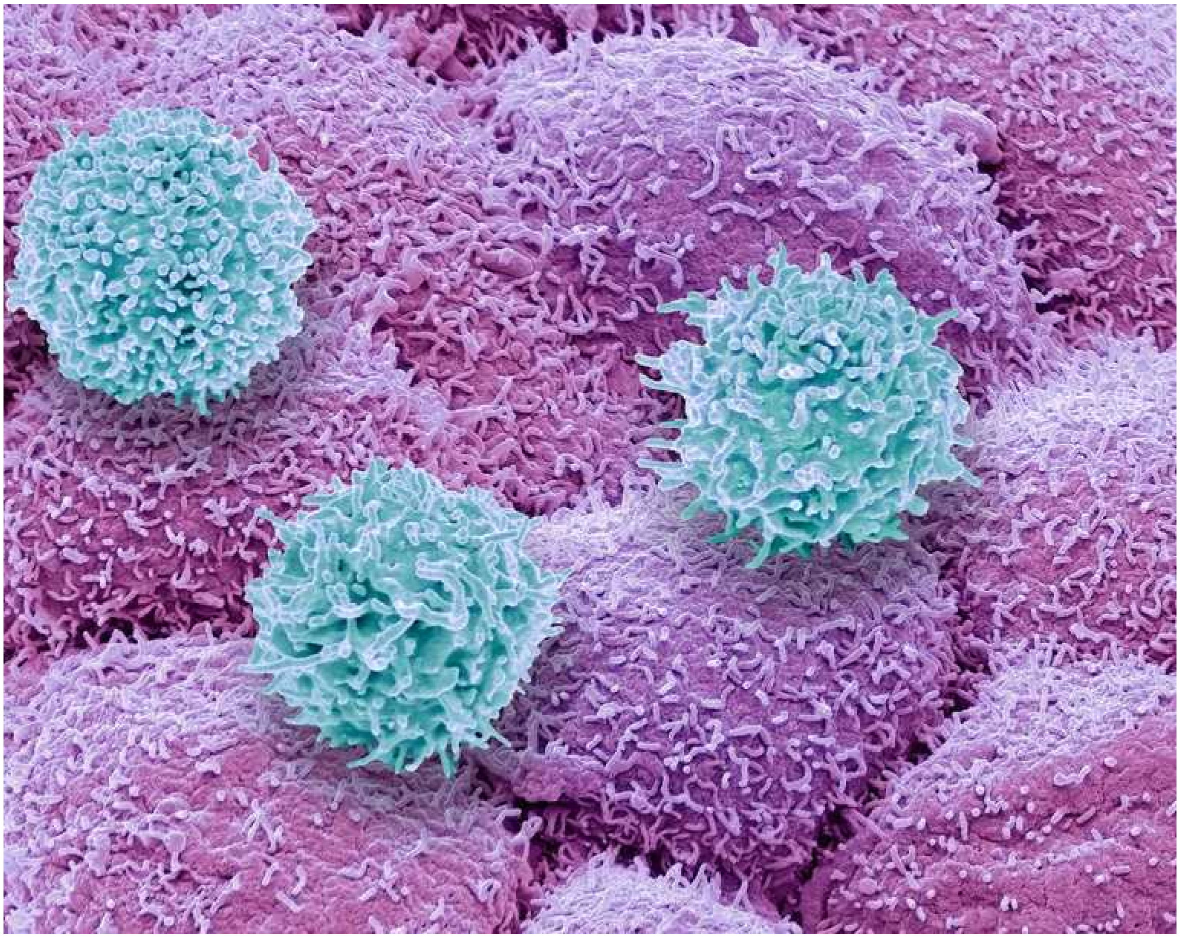
Revved up and ready to go

CAR-T therapies capitalize on the activities of T cells, the immune system’s natural hunters that prowl through the body looking for things that don’t belong. Foreign cells, or those infected with a virus, express unusual proteins that serve as a beacon to T cells, some of which release a toxic stew of molecules to destroy the abnormal cells. This search-and-destroy function can also target cancer cells for elimination, but tumours often have ways of disarming the immune system, such as by cloaking abnormal proteins or suppressing T-cell function.

CAR-T cells carry synthetic proteins — the chimeric antigen receptors — that span the cell membrane. On the outside is a structure that functions like an antibody, binding to specific molecules on the surface of some cancer cells. Once that has bound, the portion of the protein inside the cell stimulates T-cell activity, hot-wiring it into action. The result is a tiny, revved-up, cancer-fighting machine.

Approved CAR-T therapies target one of two proteins found on immune cells called B cells, and are used to treat certain forms of leukaemia and lymphoma that involve the unchecked proliferation of these cells. The proteins — CD19 and BCMA — are not unique to cancer, meaning that the

therapies kill B cells indiscriminately. However, people can live without these cells.



T cells (blue) of the immune system attacking prostate cancer cells (pink).Credit: Steve Gschmeissner/SPL

There is still plenty of room for improvement in CAR-T therapies. Although the effects can be long-lasting — sometimes even curative — cancer eventually returns in most people who have been treated. Solid tumours, such as those found in lung or pancreatic cancers, have so far not responded convincingly to CAR-T cells. The therapy has safety risks and can, in rare instances, be fatal. And it must be custom-made for each recipient, using their own T cells as a starting point, resulting in a relatively slow and expensive manufacturing process.

As yet, there are no simple solutions to any of these problems. “We clearly have a long way to go,” says Mackall. “But we’re now seeing promising

signals."

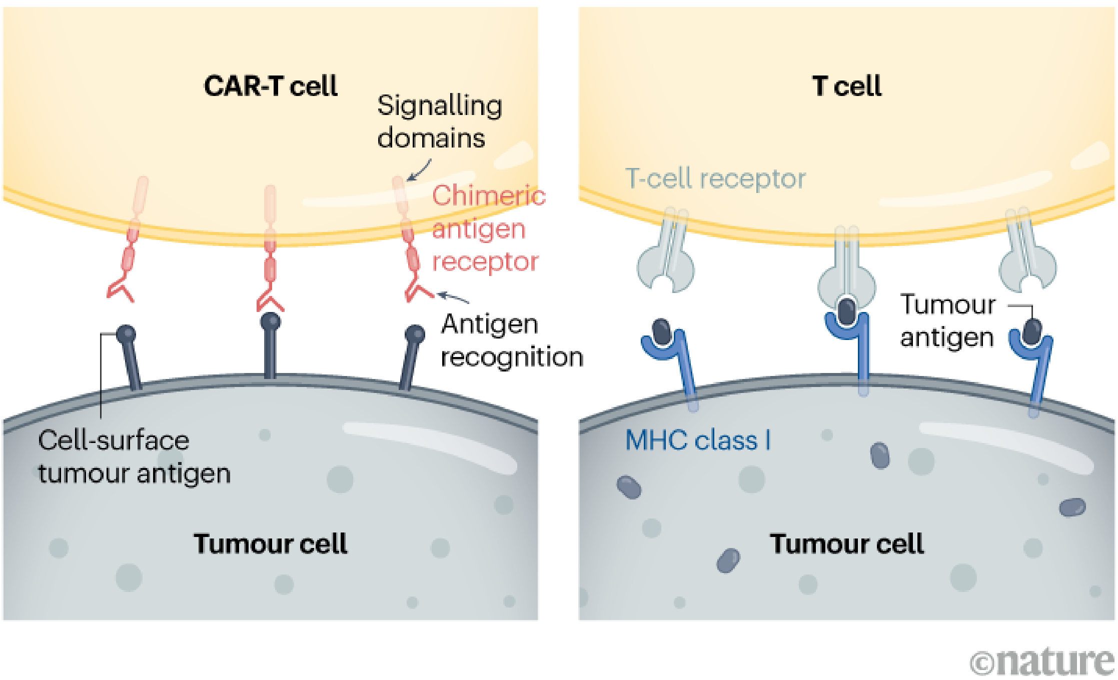
Some progress is being made against solid tumours. These often contain a heterogeneous mosaic of cells that have different combinations of mutations. This means that a CAR-T therapy directed at a particular mutated protein might work for only one subset of cells. The tight mass of a solid tumour can also be difficult for T cells to penetrate, and researchers have struggled to find suitable targets that won't wreak havoc in healthy tissues.

Despite this, some clinical trials have shown glimmers of efficacy. Mackall and her colleagues have engineered CAR-T cells to target a molecule called GD2, which is expressed at high levels by some brain and spinal-cord cancers called gliomas. The team gave one intravenous dose of CAR-T therapy to people with gliomas, then administered multiple, lower doses directly into the brain. She and her colleagues reported last year that three of four people treated in this way responded positively[1]. "These cells just dive right into the brain," says Mackall. "And the body doesn't reject them up there — it's playing in that immune-privileged space."

Targeting solid tumours could require T-cell therapies that recognize more than one mutated protein or that can target cancer cells expressing higher levels of a given protein than normal cells do. One clinical trial that reported results in November 2022 took this to the extreme: rather than using CARs, the team used CRISPR to engineer natural T-cell receptors (see 'Targeting T cells') to recognize mutated proteins found in each participant's tumour[2]. The individuals received a mixture of cells targeting different proteins, in the hope that solid tumours would be less likely to develop resistance to a therapy with multiple targets. Tumours stopped growing in 5 of the 16 participants 28 days after treatment. Researchers hope to tweak the protocol, including giving higher doses, to boost effectiveness.

TARGETING T CELLS

A promising form of cancer treatment uses modified T cells — part of the immune system — to kill tumours. Most approved therapies (left panel) contain a chimeric antigen receptor (CAR): an artificial, engineered receptor that recognizes a protein, or antigen, on the surface of a cancer cell and signals the T cell to attack. Other approaches (right panel) have re-engineered the T cell's natural receptors. This allows T cells to target antigens not normally on the surface of a cancer cell if they are displayed by the protein MHC class I, a component of the immune system.



Source: Premier Research; adapted from https://go.nature.com/3WXCRYX

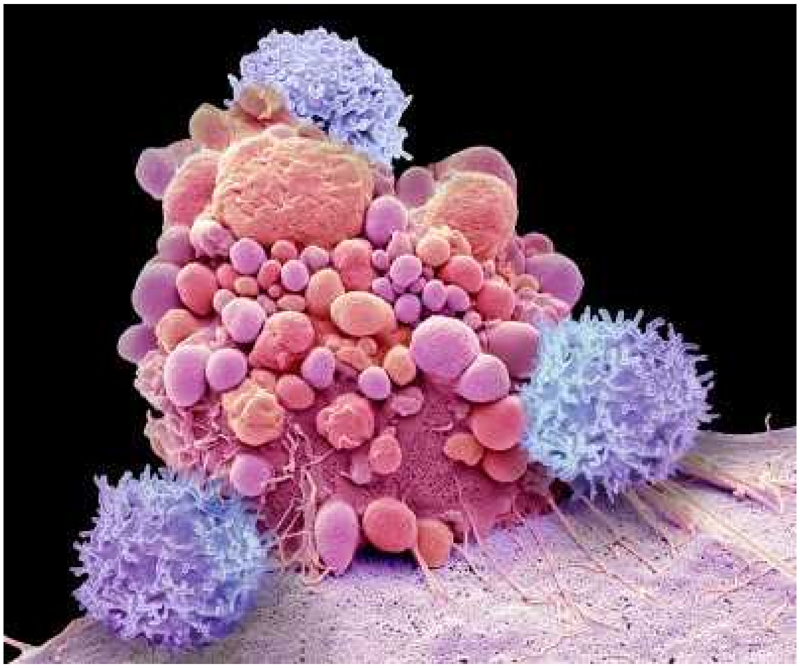
The ability to track and fine-tune T-cell activity is also improving, says immunologist Carl June at the University of Pennsylvania. Through advanced single-cell analyses, researchers can follow the fate of both the engineered cells and the tumours they are designed to kill. They can determine which T cells have become 'exhausted' — a dysfunctional state that can come from prolonged stimulation — and which tumour cells are becoming resistant to treatment. They can also see whether the environment surrounding a CAR-T-treated tumour has become riddled with immune-suppressing cells (such as macrophages or regulatory T cells). Overcoming that local immune suppression will be key to harnessing T cells to fight solid tumours, says Yangbing Zhao, chief scientific officer at UTC Therapeutics, a biotechnology company headquartered in Singapore that is developing

CAR-T therapies. “No matter how many targets you target, if the tumour is evading the immune response, it won’t work,” he says.

June and his colleagues used a single-cell approach to study resistance to CAR-T therapies that target CD19, and found that CAR-T products that were less able to activate certain helper T cells were associated with the emergence of resistance[3]. They also used single-cell techniques to learn more about why CAR-T cells directed against a protein called mesothelin, found in pancreatic cancer cells, often fail. Reducing the activity of two genes in CAR-T cells might bolster the therapy[4]. “We’re going to be able to understand these resistance mechanisms,” says June. “And then with all of these tools like CRISPR, we’re going to engineer around them.”

In addition to editing T cells, CRISPR has been used to find more ways of modifying them. Immunologist Alexander Marson at the Gladstone Institutes in San Francisco, California, and his colleagues used CRISPR to activate or suppress thousands of genes in T cells, and then looked at the effect the changes had on the production of crucial immune-regulating proteins called cytokines[5]. In another screen using CRISPR, the team found that reducing the activity of a protein called RASA2 enhanced the ability of CAR-T cells to kill their targets[6]. “We’re learning lessons about the genes that we can turn up and turn down to tune T cells to behave as we want,” says Marson.

Synthetic biologists have also set their sights on T cells, and are engineering sophisticated cellular circuits that could allow greater control over the expression of CARs and other proteins that might increase T-cell activity. In December last year, synthetic biologist Wendell Lim at the University of California, San Francisco, and his colleagues reported[7] that they had engineered T cells to express both a CAR and IL-2, an immune-regulating protein. IL-2 can improve T-cell penetration into solid tumours and overcome the immunosuppressive signals that tumours release, but it can be toxic when administered systemically. Letting the T cells produce IL-2 enables local administration of the protein, which could bypass its toxicity to other tissues.



Last-resort cancer therapy holds back disease for more than a decade

Other synthetic circuits have been designed to allow precise regulation of CAR expression, by placing it under the control of genetic elements that activate the necessary genes in response to a drug[8]. So far, however, most of these complicated designs have not yet gone through the safety studies and standardization required for use in people, says Sadelain.

Researchers are learning so many lessons that a big question for the field is now determining which engineered T cells to take forwards into human studies, says oncologist Marcela Maus at Massachusetts General Hospital in Boston. “We can invent and innovate so much in the lab, but there is this funnel of translating that into clinical trials,” she says. “There’s so many things we can do. We have to figure out which are the best things to tweak and test in trials.”

Costly business

Manufacturing CAR-T cells is already wildly complex by pharmaceutical standards. So far, all approved therapies require engineering a person’s own T cells to express the CAR. That adds to the time and thus the cost of producing the therapies: in the United States, a single treatment with CAR-T

cells can be about US$500,000, not including the cost of hospitalization and associated treatments.

Creating CAR-T cells that can be given to multiple people — often called off-the-shelf cells — has long been viewed as crucial to lowering the price of the therapy. But early results suggest that there is still work to do, says bioengineer Rahul Purwar at the Indian Institute of Technology Bombay. Although the cells can be edited to reduce the chance that they will themselves be eliminated by the immune system, early trials suggest that they do not survive long after infusion and might still be rejected (see, for example, ref. 9)[9]. "Off-the-shelf is a great approach," he says. "It is coming, but right now we are not yet there."



Cancer treatments boosted by immune-cell hacking

The therapy is also rarely available outside wealthy countries. In Brazil, haematologist Renato Luiz Guerino Cunha at Oncoclínicas Group in São Paulo was the first in the country to treat someone with CAR-T therapy in 2019. But progress has been slow, he says: he lacks the capacity to rapidly produce large quantities of cells. "In three years, we treated just six patients," he says. "We need new technology for the processing."

Producing a CAR-T cell therapy typically involves using a type of virus called a lentivirus as a vector to shuttle in the synthetic CAR gene. But more

research into gene therapies has increased demand for clinical-grade lentiviruses. Researchers now wait months and pay top dollar to complete their experiments; Cunha produces his own but can do so only in tiny quantities. Improvements to CRISPR gene editing could help in this regard.

Despite the challenges, CAR-T therapies continue to expand, with some of the hundreds of clinical trials worldwide exploring entirely new applications. Last year, researchers reported promising results in a small trial of CAR-T therapies to treat a form of the autoimmune disease lupus[10]. And in a study in mice, researchers reprogrammed T cells without the usual first step of removing them from the body, creating CAR-T cells designed to clear scar tissue from the heart[11].

In December, June and his colleagues unveiled a way to streamline cell production. At the American Society of Hematology's annual meeting in New Orleans, Louisiana, the team announced[12] that reducing manufacturing times and engineering CAR-T cells to express a protein called IL-18 boosted their efficacy and allowed researchers to reduce the dose of cells given to people. "Those patients had incredible responses," says Maus of the clinical trial, "which gives you this really tantalizing hint that if you engineer the T cell better, you can make it even more powerful."

Nature **613**, 626-628 (2023)

doi: https://doi.org/10.1038/d41586-023-00177-z

Updates & Corrections

- **Correction 27 January 2023**: An earlier version of this feature incorrectly referred to GD2 as a protein.

References

1. Majzner, R. G. *et al. Nature* **603**, 934–941 (2022).

2. Foy, S. P. *et al. Nature* https://doi.org/10.1038/s41586-022-05531-1 (2022).

3. Bai, Z. *et al. Sci. Adv.* **8**, eabj2820 (2022).

4. Good, C. R. *et al. Cell* **184**, 6081–6100 (2021).

5. Schmidt, R. *et al. Science* **375**, eabj4008 (2022).

6. Carnevale, J. *et al. Nature* **609**, 174–182 (2022).

7. Allen, G. M. *et al. Science* **378**, eaba1624 (2022).

8. Li, H-S. *et al. Science* **378**, 1227–1234 (2022).

9. Benjamin, R. *et al. Lancet Haematol.* **9**, E833–E843 (2022).

10. Mackensen, A. *et al. Nature Med.* **28**, 2124–2132 (2022).

11. Rurik, J. G. *et al. Science* **375**, 91–96 (2022).

12. Svoboda, J. *et al*. *Blood* **140** (Suppl. 1), 4612–4614 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00177-z

| Section menu | Main menu |

Opinion

Work

- **How philanthropy can nurture your research** [13 Jan 2023]
  • Scientists deploy networking, collaboration and long-term thinking to win private donor dollars.
- **Seven technologies to watch in 2023** [23 Jan 2023]
  • Nature’s pick of tools and techniques that are poised to have an outsized impact on science in the coming year.
- **Why I’m a garbage collector for the world’s oceans** [23 Jan 2023]
  • Boyan Slat founded and operates a business that seeks to reduce ocean plastic pollution by 90%.

- CAREER FEATURE
- 13 January 2023
- Correction 17 January 2023

How philanthropy can nurture your research

Scientists deploy networking, collaboration and long-term thinking to win private donor dollars.

- Emily Sohn

Find a new job



Credit: Getty

It takes money to do research and, increasingly, much of it comes from private donors and foundations, especially in the United States. Philanthropists provided 44% of research funding at US universities and non-profit research institutes in 2016, according to a 2018 post by the Science Philanthropy Alliance. This collective of philanthropists and foundations, which aims to build privately funded support for basic science in the United States, the United Kingdom and Canada, has grown from 6 members since it launched in 2013 to more than 30.



Nature Collection: Funding science

The money adds up. Overall, philanthropic groups in the United States contribute US$30 billion annually to science, similar to the amount spent by federal agencies, according to a 2022 analysis of US non-profit tax returns[1]. That's more than in Europe, where a 2016 review[2] of about 1,000 foundations estimated that philanthropies contributed more than $6.4 billion a year to research and innovation. That is only around 2% of the $353.6 billion total for the region, which was provided mostly by governments and industry. Many of these philanthropic grants are open to international researchers.

The landscape of philanthropic funding differs from publicly funded grants in key ways. Among them is the fact that there are many thousands of organizations each doling out relatively small amounts (see 'Modest donations add up').

Modest donations add up

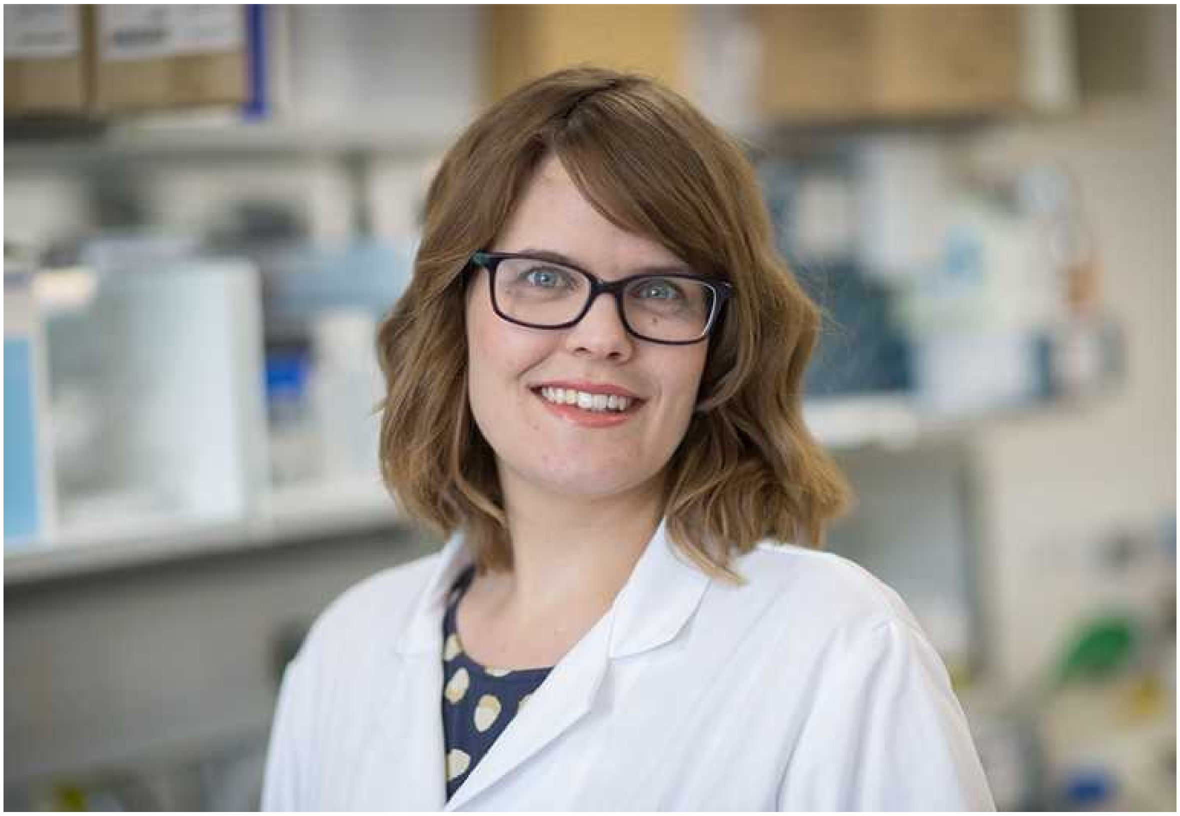
To quantify philanthropic giving to institutions involved in science, a 2022 study[1] looked at the tax returns of 685,397 US non-profit organizations in 2010–19. The authors found that:

• The largest 200 donors accounted for 66% of the philanthropic money given to US science — but those big donors represented only 0.3% of organizations that give out grants.

• Around 13,000 science-related organizations received philanthropic support, compared with the 15,000 supported by the US National Institutes of Health (NIH) and 10,000 supported by NSF.

• Philanthropic funding was distributed more widely than federal funds. Whereas 200 organizations received 80% of NIH and NSF grants, the top 200 recipients of philanthropic support got around 33% of all the donated money.

Because of the wide range of grants available, it can be challenging to find out about the various opportunities and to keep track of them, researchers say. Winning such grants also requires different approaches. *Nature* spoke to four scientists about how they have acquired private donations, and what that funding has enabled them to accomplish.



Immunologist Danika Hill won the Michelson Prize in 2020, which has powered her research on the immune response to infection.Credit: Babraham Institute

DANIKA HILL: Embrace risk

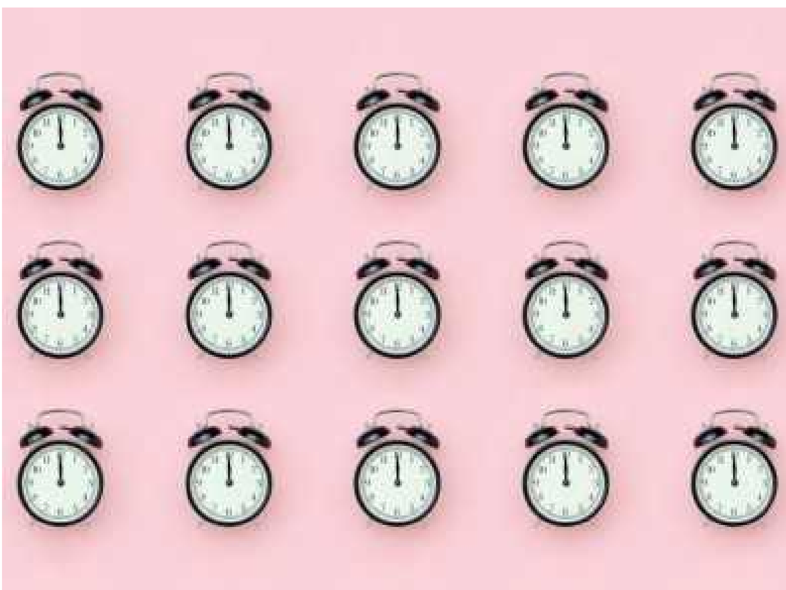
Immunologist at Monash University in Melbourne, Australia.

To become an independent researcher, you must prove that you can get funding, but to compete for most national funding bodies, you have to already be independent.

I started looking around for philanthropic funding during my postdoc at the University of Cambridge, UK, because I was looking to move back to Australia and to transition to independence. As a postdoc, I wasn't eligible for any of the big pots of money from UK national funders, such as the Biotechnology and Biological Sciences Research Council and the Medical Research Council. It's like trying to find a doorway that you then can't get through.

I found out about the Michelson Prize, which aims to support research that accelerates solutions to global health challenges, through my group leader, who had said, “Keep this on your radar.” I had to apply before I turned 35. I regularly checked the website to see when the call would be open. Often, these funding calls go from no announcement to, “You’ve got a month.” You have to be ready to pounce.

There’s no central repository where you can find out about all the possible private funding avenues. A lot of schemes can be quite small and tailored, so mapping the landscape is really challenging. You have to either do your research or have people alerting you to these possibilities. I asked people who are further on in their careers whether they knew of any funding, fellowships or prizes that I should be aware of. I got lots of helpful advice. Otherwise, it requires trawling the Internet.



Seven questions to ask before pursuing a new grant

For the Michelson Prize, the committee was looking for high-risk, high-reward projects that were going to change the field. That type of project wouldn’t be fundable through conventional grant schemes, for which you must do the risk-mitigation beforehand. A nice thing about the philanthropic funding was they just wanted to invest in an idea that hadn’t been done before.

My project involved tackling vaccine design from a new perspective. I proposed finding the immune cells involved in generating a long-lived immune response to infection, and then working backwards to find out which parts of a pathogen these cells recognized. I hope this approach will increase the chances of long-lived immunity, but it has never been done before. In my application, I made sure to emphasize the novelty, while highlighting that I had the skills necessary to execute the project.

Each funder will have a different appetite for risk, and different priorities for funding basic science, early-stage research or something close to translation to the clinic. Tailor your proposal to appeal to their specific agenda.

Since I got the prize in 2020, the Michelson Medical Research Foundation, which administers the awards, has used me in advertising materials on social media. I think it has done wonders for my reputation. I can now show that I was one of only two people who won this international funding call. It's about reputation, prestige and exposure. I gave a 15-minute talk as part of the award ceremony; afterwards, I was contacted by a top-tier journal editor who said, "Your science is really cool. Let me know when you're ready to publish." That doesn't happen often as an early-career researcher. I've had potential collaborators reach out and say, "Why don't we work together?" I also got headhunted for a possible position at another institute, showing that I was on the radar of bigger names in my field.

In my experience, philanthropic funding is often quite narrow in the scope of what and who the funder is looking for. So, you should tailor your research to match their priorities and focus areas. Often, they don't cover the full cost of research, so you need a supportive institution or group leader behind you to cover that shortfall. They can have quite strict timelines, so you have to make sure that you've got the time and resources to actually do the research in the time frame.

My work is still a way off from publishing, but I've got lots of data. Now I can go to those national funding bodies, and not only can I show that this international organization thought this was a cool idea, but I've also done a lot of the research, and it's working. I can say, "Wouldn't it be great if you gave me some more money to push it over the line?" It's gone a huge way to

show that I'm fundable, and that I'm really starting my own, independent research direction.



Physicist Rafael Chaves says philanthropic support makes up the majority of his research funding.Credit: Cyro Lucas Souza

RAFAEL CHAVES: Put your name out there

Physicist in the International Institute of Physics at the Federal University of Rio Grande do Norte in Natal, Brazil.

All my major grants so far are from philanthropic institutions. I received the first one in 2015, during my postdoc at the University of Freiburg in Germany, from the Foundational Questions Institute (FQXi) in Decatur, Georgia. It was not much, around $100,000, but it was enough to organize two or three conferences and to visit collaborators. I was working on a topic that was relatively new, and we did the first series of events in quantum causality and quantum networks. It was a great opportunity to meet more

people and become known. Travelling and making connections has been key to my success.

I applied for the FQXi grant with another junior researcher and a senior one. Having someone with a big name made things easier. It shouldn't be like this, but the support of a senior researcher can make things easier for the next application. I feel it's an important first step to get some kind of grant to start, even if it's not the biggest or most renowned one. It shows you can manage research projects of your own.



The hunt for the lesser-known funding source

In 2016, I started my current permanent position as a research leader at the International Institute of Physics in Natal, Brazil. Just after I arrived, I applied for a John Templeton Foundation grant with a collaborator, Fabio Sciarrino, and we got it. (The Templeton Foundation in West Conshohocken, Pennsylvania, funds research in physics, biology, philosophy, religion, free markets and culture.) It was around $200,000. It allowed me to hire a postdoc, to organize a high-profile 2019 conference on causality in the quantum world in Capri, Italy, and to travel to conferences and visit collaborators. It was a game-changer, because Brazil is far away from the big research centres and travelling is always expensive.

I also got a grant in 2019 from Serrapilheira, a Brazilian institute that launched in 2017 with a 350-million reais (US$64-million) endowment fund. As far as I know, it's the only philanthropic institute that funds science in Brazil. Its budget is considerably bigger than what you get from the government here, and is comparable with European and North American grants. The institute helped me to start doing science outreach, and in September 2022, with its support, I published my first book about quantum information for the public, called *Incerteza Quântica* (which means quantum uncertainty in Portuguese).

With my philanthropic grants, if some issue came up, I could make a phone call to say, "I'm facing this problem. Could I rearrange the budget a little bit?" But with a government grant, that kind of conversation is basically impossible.

The science budget in Brazil was cut by more than half during Jair Bolsonaro's administration. The academic community is really hoping for that to turn around, and is very supportive of the newly elected president, Luiz Inácio Lula da Silva. Without philanthropic grants, it would be almost impossible for me to do any science, hire postdocs, travel for conferences or invite people here. It would just be teaching at the university and trying to keep up with science at the frontier, while being isolated in my beautiful tropical country.



Molecular neuroscientist Hermona Soreq often speaks to donors on behalf of her institution, The Hebrew University of Jerusalem.Credit: David Salem/ZOOG PRODUCTIONS

HERMONA SOREQ: Befriend donors

Molecular neuroscientist at The Hebrew University of Jerusalem.

In Israel, science is not as well funded by the government as scientists would hope compared to other spending. As scientists there, we spend a lot of time applying for grants and thinking about the next one, and it doesn't get easier with time because we live in a country where there's always a crisis that needs funding for the military. Science is always at the end of a long list of priorities, which is a real problem.

My links with philanthropic organizations are mainly to support my university, which I see as part of my job. The Hebrew University has an organization of donors known as Friends, and they support the university very generously. We largely depend on their support to fund fellowships for

research students, to provide start-up equipment for new faculty members and to support core facilities. They are interested in what the university is doing, and they become personal friends. I have several, and I keep e-mailing and talking to them about my research. They come to visit, and keep supporting us.

The university often asks me to meet with Friends when I am in the United States. They always have issues that they would like to see studied. I find relevant experts at our centre, institute or university, and connect them to these potential funders.

In 2020 I delivered a lecture at the Sephardic Foundation on Aging, a non-profit organization in New York City that focuses on neurodegenerative diseases, which is linked to the brain circuit that I'm interested in. They said, 'We want to find out why Jewish people develop Alzheimer's disease,' so I connected them with a colleague who studies Alzheimer's. And the approach works — they support fellowships for research students at our university.

I'm the senior scientific adviser for the Azrieli Foundation's fellows programme, which provides fellowships to PhD students, postdocs and other junior scientists in Israel. It's always interesting to judge for those committees. You get to change someone's life for the better in a way that I like, because I love science. Candidates who combine expertise from different disciplines stand out as special. My advice is: keep working and try to be the best, because it pays.



Physicist Steven Kahn approached big donors to kick-start funding for the Vera C. Rubin Observatory project that he directs.Credit: SLAC National Accelerator Laboratory

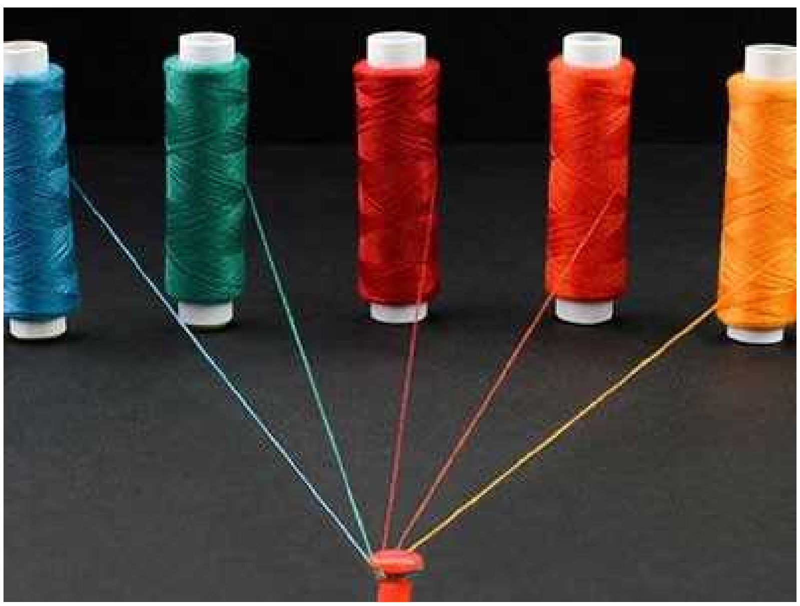
STEVEN KAHN: Seek leverage

Physicist and dean of mathematical and physical sciences at the University of California, Berkeley.

In 2013, I became director of the Large Synoptic Survey Telescope, which was renamed the Vera C. Rubin Observatory in 2019. Work started on the project in earnest around 2003. By 2009, we had acquired private funding to kick-start the project, and US federal agencies approved it in 2014. The observatory should be completed, commissioned and ready to begin regular operations in July 2024. It will take an image of the entire southern sky every 15 seconds every night for 10 years. By adding all the images together, the resulting data set can be used to measure everything that moves and changes brightness in the sky.

It’s a $670-million construction project in the Coquimbo region of northern Chile, and is now almost totally funded by the US National Science Foundation and the Department of Energy. However, with federal funds, you can’t spend money ahead of time on fabricating any components until the project is formally approved. So we needed to get private donors involved.

For us, the item that would determine the pace of the whole project was the telescope’s primary mirror, which is a very large, precision piece of glass. It was a catch-22 situation, because once we got approval, we were going to need to spend five years trying to make the mirror. The argument we made to donors was that if we could get private funding to enable us to fabricate that mirror earlier, that would get the project kick-started, and once approved, that would allow us to build the observatory.



Getting the job: it’s not just who you know, but how you know them

The largest private gifts came in around 2008 and 2009. We collected $20 million from the Charles and Lisa Simonyi Fund for Arts and Sciences, and $10 million from Microsoft co-founder Bill Gates personally. We got an additional $10 million from various other donors, ending up with about $40 million in private money that enabled us to start on the mirror. We parlayed that into more than half a billion dollars of federal money.

The federal research awards process, from the very small grants to huge projects, is largely driven by peer review. This means that scientists need to convince other scientists that this is a good use of money, either by writing proposals or by trying to drum up support. At the individual donor level, by contrast, much is face to face — a lot of it is personal chemistry. Academics often get this wrong, because if you talk to donors the way you would talk to colleagues, they don't always follow. Yet if you talk to them like they're the general public, you can over-explain and it becomes a little bit patronizing. Try to strike a balance.

Donors need to hear an argument that makes sense, not just a bunch of feel-good stuff. When donors go into these meetings, they want to have an idea of why the money is important, what we would actually do with it and how it will make a difference. Don't over-promise and say, 'We're going to find a cure for cancer.' Instead, you can say $100,000 will help to support a couple of graduate students. Your message must make sense in the context of the gift that you're requesting.

The odds of winning funding from any encounter are not that high. You can't go into it thinking, 'I'm going to give a talk at this dinner and get a big reward out of it.' You have to think 'it's good to maintain a relationship with alumni or potential donors' — and that eventually, someday, this might pay off.

Someone beginning an assistant professorship at a university, or a similar role, should focus on getting federal and other publicly funded awards to support their research. But they should also make themselves more knowledgeable on the philanthropic side. Universities have directors of development and related staff who can connect interested donors to relevant researchers. In my experience, donors love lunch with someone who's a young researcher doing exciting things. It's a good idea to be thinking about this early in your career.

Nature **613**, 791-793 (2023)

doi: https://doi.org/10.1038/d41586-023-00077-2

These interviews have been edited for length and clarity.

Updates & Corrections

- **Correction 17 January 2023**: An earlier version of this feature erroneously stated that Hermona Soreq is an adviser for the Azrieli Foundation. In fact, she is an adviser for the foundation's fellows programme.

References

1. Shekhtman, L. M. *et al.* Preprint at https://arxiv.org/abs/2206.10661 (2022).

2. Gouwenberg, B. *et al. Found. Rev*. https://doi.org/10.9707/1944-5660.1287 (2016).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00077-2

- TECHNOLOGY FEATURE
- 23 January 2023

Seven technologies to watch in 2023

Nature's pick of tools and techniques that are poised to have an outsized impact on science in the coming year.

- Michael Eisenstein



The James Webb Space Telescope's 6.5-metre primary mirror (6 of 18 segments shown) can detect objects billions of light years away. Credit: NASA/MSFC/David Higginbotham

From protein sequencing to electron microscopy, and from archaeology to astronomy, here are seven technologies that are likely to shake up science in

the year ahead.

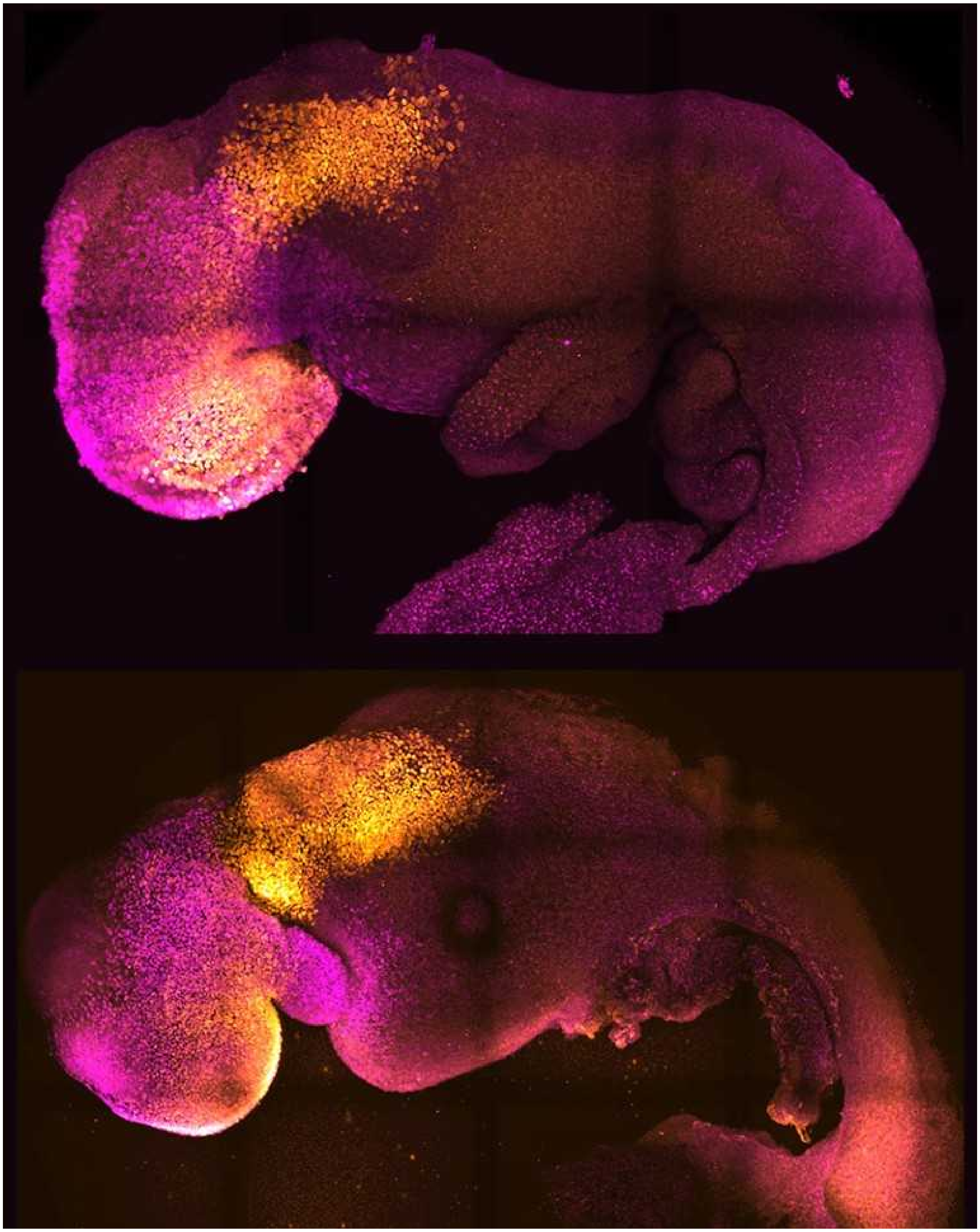
Single-molecule protein sequencing

The proteome represents the complete set of proteins made by a cell or organism, and can be deeply informative about health and disease, but it remains challenging to characterize.

Proteins are assembled from a larger alphabet of building blocks relative to nucleic acids, with roughly 20 naturally occurring amino acids (compared with the four nucleotides that form molecules such as DNA and messenger RNA); this results in much greater chemical diversity. Some are present in the cell as just a few molecules — and, unlike nucleic acids, proteins cannot be amplified, meaning protein-analysis methods must work with whatever material is available.

Most proteomic analyses use mass spectrometry, a technique that profiles mixtures of proteins on the basis of their mass and charge. These profiles can quantify thousands of proteins simultaneously, but the molecules detected cannot always be identified unambiguously, and low-abundance proteins in a mixture are often overlooked. Now, single-molecule technologies that can sequence many, if not all, of the proteins in a sample could be on the horizon — many of them analogous to the techniques used for DNA.

Edward Marcotte, a biochemist at the University of Texas at Austin, is pursuing one such approach, known as fluorosequencing[1]. Marcotte's technique, reported in 2018, is based on a stepwise chemical process in which individual amino acids are fluorescently labelled and then sheared off one by one from the end of a surface-coupled protein as a camera captures the resulting fluorescent signal. "We could label the proteins with different fluorescent dyes and then watch molecule by molecule as we cut them away," Marcotte explains. Last year, researchers at Quantum-Si, a biotechnology firm in Guilford, Connecticut, described an alternative to fluorosequencing that uses fluorescently labelled 'binder' proteins to recognize specific sequences of amino acids (or polypeptides) at the ends of proteins[2].



Researchers can now make synthetic embryos in the laboratory (bottom) that resemble natural, eight-day-old embryos (top).Credit: Magdalena Zernicka-Goetz Laboratory

Other researchers are developing techniques that emulate nanopore-based DNA sequencing, profiling polypeptides on the basis of the changes they induce in an electric current as they pass through tiny channels. Biophysicist Cees Dekker at Delft University of Technology in the Netherlands and his colleagues demonstrated one such approach in 2021 using nanopores made of protein, and were able to discriminate between individual amino acids in a polypeptide passing through the pore[3]. And at the Technion — Israel Institute of Technology in Haifa, biomedical engineer Amit Meller's team is investigating solid-state nanopore devices manufactured from silicon-based materials that could enable high-throughput analyses of many individual protein molecules at once. "You might be able to look at maybe tens of thousands or even millions of nanopores simultaneously," he says.

Although single-molecule protein sequencing is only a proof of concept at present, commercialization is coming fast. Quantum-Si has announced plans to ship first-generation instruments this year, for example, and Meller notes that a protein-sequencing conference in Delft in November 2022 featured a discussion panel dedicated to start-ups in this space. "It reminds me a lot of the early days before next-generation DNA sequencing," he says.

Marcotte, who co-founded the protein- sequencing company Erisyon in Austin, Texas, is bullish. "It's not really a question of whether it will work," he says, "but how soon it will be in people's hands."

James Webb Space Telescope

Astronomers began last year on the edges of their collective seats. After a design and construction process lasting more than two decades, NASA — in collaboration with the European and Canadian space agencies — successfully launched the James Webb Space Telescope (JWST) into orbit on 25 December 2021. The world had to stand by for nearly seven months as the instrument unfolded and oriented itself for its first round of observations.

It was worth the wait. Matt Mountain, an astronomer at the Space Telescope Science Institute in Baltimore, Maryland, who is a telescope scientist for JWST, says the initial images exceeded his lofty expectations. "There's actually no empty sky — it's just galaxies everywhere," he says.

“Theoretically, we knew it, but to see it, the emotional impact is very different.”

JWST was designed to pick up where the Hubble Space Telescope left off. Hubble generated stunning views of the Universe, but had blind spots: ancient stars and galaxies with light signatures in the infrared range were essentially invisible to it. Rectifying that required an instrument with the sensitivity to detect incredibly faint infrared signals originating billions of light years away.

The final design for JWST incorporates an array of 18 perfectly smooth beryllium mirrors that, when fully unfolded, has a diameter of 6.5 metres. So precisely engineered are those mirrors, says Mountain, that “if you stretched a segment out over the United States, no bump could be more than a couple of inches [high].” These are coupled with state-of-the-art near- and mid-infrared detectors.



NatureTech

That design allows JWST to fill in Hubble’s gaps, including capturing signatures from a 13.5-billion-year-old galaxy that produced some of the first atoms of oxygen and neon in the Universe. The telescope has also yielded some surprises; for instance, being able to measure the atmospheric composition of certain classes of exoplanet.

Researchers around the world are queueing up for observation time. Mikako Matsuura, an astrophysicist at Cardiff University, UK, is running two studies with JWST, looking at the creation and destruction of the cosmic dust that can contribute to star and planet formation. “It’s a completely different order of sensitivity and sharpness” compared with the telescopes her group has used in the past, Matsuura says. “We have seen completely different phenomena ongoing inside these objects — it’s amazing.”

Volume electron microscopy

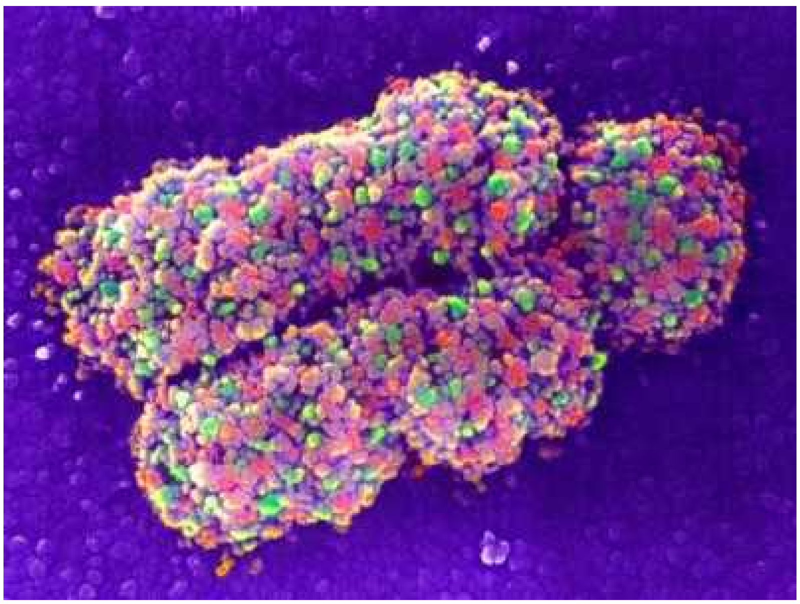
Electron microscopy (EM) is known for its outstanding resolution, but mostly at the surface level of samples. Going deeper requires carving a specimen into exceptionally thin slices, which for biologists are often insufficient for the task. Lucy Collinson, an electron microscopist at the Francis Crick Institute in London, explains that it can take 200 sections to cover the volume of just a single cell. “If you’re just getting one [section], you’re playing a game of statistics,” she says.

Now researchers are bringing EM resolution to 3D tissue samples encompassing many cubic millimetres.

Previously, reconstructing such volumes from 2D EM images — for example, to chart the neural connectivity of the brain — involved a painstaking process of sample preparation, imaging and computation to turn those images into a multi-image stack. The latest ‘volume EM’ techniques now drastically streamline this process.

Those techniques have various advantages and limitations. Serial block-face imaging, which uses a diamond-edged blade to shave off thin sequential layers of a resin-embedded sample as it is imaged, is relatively fast and can handle samples approaching one cubic millimetre in size. However, it offers poor depth resolution, meaning the resulting volume reconstruction will be comparatively fuzzy. Focused ion beam scanning electron microscopy (FIB-SEM) yields much thinner layers — and thus finer depth resolution — but is better suited to smaller-volume samples.

Collinson describes the rise of volume EM as a ‘quiet revolution’, with researchers highlighting the results of this approach rather than the techniques used to generate them. But this is changing. For example, in 2021, researchers working on the Cell Organelle Segmentation in Electron Microscopy (COSEM) initiative at Janelia Research Campus in Ashburn, Virginia, published a pair of papers in *Nature* highlighting substantial progress in mapping the cellular interior[4,5]. “It’s a very impressive proof of principle,” says Collinson.



Seven technologies to watch in 2022

The COSEM initiative uses sophisticated, bespoke FIB-SEM microscopes that increase the volume that can be imaged in a single experiment by roughly 200-fold, while preserving good spatial resolution. Using a bank of these machines in conjunction with deep-learning algorithms, the team was able to define various organelles and other subcellular structures in the full 3D volume of a wide range of cell types.

The sample-preparation methods are laborious and difficult to master, and the resulting data sets are massive. But the effort is worthwhile: Collinson is already seeing insights in infectious-disease research and cancer biology. She is now working with colleagues to explore the feasibility of reconstructing the entire mouse brain at high resolution — an effort she predicts will take more than a decade of work, cost billions of dollars and

produce half a billion gigabytes of data. “It’s probably on the same order of magnitude as the effort to map the first human genome,” she says.

CRISPR anywhere

The genome-editing tool CRISPR–Cas9 has justifiably earned a reputation as the go-to method for introducing defined changes at targeted sites throughout the genome, driving breakthroughs in gene therapy, disease modelling and other areas of research. But there are limits as to where it can be used. Now, researchers are finding ways to circumvent those limitations.

CRISPR editing is coordinated by a short guide RNA, which directs an associated Cas nuclease enzyme to its target genomic sequence. But this enzyme also requires a nearby sequence called a protospacer adjacent motif (PAM); without one, editing is likely to fail.

At the Massachusetts General Hospital in Boston, genome engineer Benjamin Kleinstiver has used protein engineering to create ‘near-PAMless’ Cas variants of the commonly used Cas9 enzyme from the bacterium *Streptococcus pyogenes*. One Cas variant requires a PAM of just three consecutive nucleotide bases with an A or G nucleotide in the middle position[6]. “These enzymes now read practically the entire genome, whereas conventional CRISPR enzymes read anywhere between 1% and 10% of the genome,” says Kleinstiver.

Such less-stringent PAM requirements increase the chances of ‘off-target’ edits, but further engineering can improve their specificity. As an alternative approach, Kleinstiver’s team is engineering and testing large numbers of Cas9 variants that each exhibit high specificity for distinct PAM sequences.

There are also many naturally occurring Cas variants that remain to be discovered. In nature, the CRISPR–Cas9 system is a bacterial defence mechanism against viral infection, and different microorganisms have evolved various enzymes with distinct PAM preferences. Virologist Anna Cereseto and microbiome researcher Nicola Segata at the University of Trento in Italy have combed through more than one million microbial genomes to identify and characterize a diverse set of Cas9 variants, which

they estimate could collectively target more than 98% of known disease-causing mutations in humans[7].

Only a handful of these will work in mammalian cells, however. “Our idea is to test many and see what are the determinants that make those enzymes work properly,” says Cereseto. Between the insights gleaned from these natural enzyme pools and high-throughput protein-engineering efforts, Kleinstiver says, “I think we’ll end with a pretty complete toolbox of editors that allow us to edit any base that we want”.

High-precision radiocarbon dating

Last year, archaeologists took advantage of advances in radiocarbon dating to home in on the precise year — and even the season — in which Viking explorers first arrived in the Americas. Working with pieces of felled timber unearthed in a settlement on the northern shore of Newfoundland, Canada, a team led by isotope-analysis expert Michael Dee at the University of Groningen in the Netherlands and his postdoc Margot Kuitems determined that the tree was likely to have been cut down in the year 1021, probably in the spring[8].

Scientists have been using radiocarbon dating of organic artefacts since the 1940s to narrow down the dates of historical events. They do so by measuring traces of the isotope carbon-14, which is formed as a result of the interaction of cosmic rays with Earth’s atmosphere and which decays slowly over millennia. But the technique is usually precise only to within a couple of decades.



Precise radiocarbon dating of timber at L’Anse aux Meadows in Newfoundland, Canada, revealed that Vikings cut down a tree at the site in 1021.All Canada Photos/Alamy

Things changed in 2012, when researchers led by physicist Fusa Miyake at Nagoya University in Japan showed[9] they could date a distinctive spike in carbon-14 levels in the rings of a Japanese cedar tree to ad 774–5. Subsequent research[10] not only confirmed that this spike was present in wood samples around the world from this period, but also identified at least five other such spikes dating as far back as 7176 bc. Researchers have linked these spikes to solar-storm activity, but this hypothesis is still being explored.

Whatever their cause, these ‘Miyake events’ allow researchers to put a precise pin in the year in which wooden artefacts were created, by detecting a specific Miyake event and then counting the rings that formed since then. Researchers can even establish the season in which a tree was harvested, on the basis of the thickness of the outermost ring, Kuitems says.

Archaeologists are now applying this approach to Neolithic settlements and sites of volcanic eruption, and Dee hopes to use it to study the Mayan empire in Mesoamerica. In the next decade or so, Dee is optimistic that “we will have really absolute records for a lot of these ancient civilizations to the exact year, and we’ll be able to talk about their historical development … at a really fine scale”.

As for Miyake, her search for historical yardsticks continues. “We are now searching for other carbon-14 spikes comparable to the 774–5 event for the past 10,000 years,” she says.

Single-cell metabolomics

Metabolomics — the study of the lipids, carbohydrates and other small molecules that drive the cell — was originally a set of methods for characterizing metabolites in a population of cells or tissues, but is now shifting to the single-cell level. Scientists could use such cellular-level data to untangle the functional complexity in vast populations of seemingly identical cells. But the transition poses daunting challenges.

The metabolome encompasses vast numbers of molecules with diverse chemical properties. Some of these are highly ephemeral, with subsecond turnover rates, says Theodore Alexandrov, a metabolomics researcher at the European Molecular Biology Laboratory in Heidelberg, Germany. And they can be hard to detect: whereas single-cell RNA sequencing can capture close to half of all the RNA molecules produced in a cell or organism (the transcriptome), most metabolic analyses cover only a tiny fraction of a cell’s metabolites. This missing information could include crucial biological insights.

“The metabolome is actually the active part of the cell,” says Jonathan Sweedler, an analytical chemist at the University of Illinois at Urbana-Champaign. “When you have a disease, if you want to know the cell state, you really want to look at the metabolites.”

Many metabolomics labs work with dissociated cells, which they trap in capillaries and analyse individually using mass spectrometry. By contrast,

‘imaging mass spectrometry’ methods capture spatial information about how cellular metabolite production varies at different sites in a sample. For instance, researchers can use a technique called matrix-assisted laser desorption/ionization (MALDI), in which a laser beam sweeps across a specially treated tissue slice, releasing metabolites for subsequent analysis by mass spectrometry. This also captures the spatial coordinates from which the metabolites originated in the sample.



Seven technologies to watch in 2021

In theory, both approaches can quantify hundreds of compounds in thousands of cells, but achieving that typically requires top-of-the-line, customized hardware costing in the million-dollar range, says Sweedler.

Now, researchers are democratizing the technology. In 2021, Alexandrov’s group described SpaceM, an open-source software tool that uses light microscopy imaging data to enable spatial metabolomic profiling of cultured cells using a standard commercial mass spectrometer[11]. “We kind of did the heavy lifting on the data-analysis part,” he says.

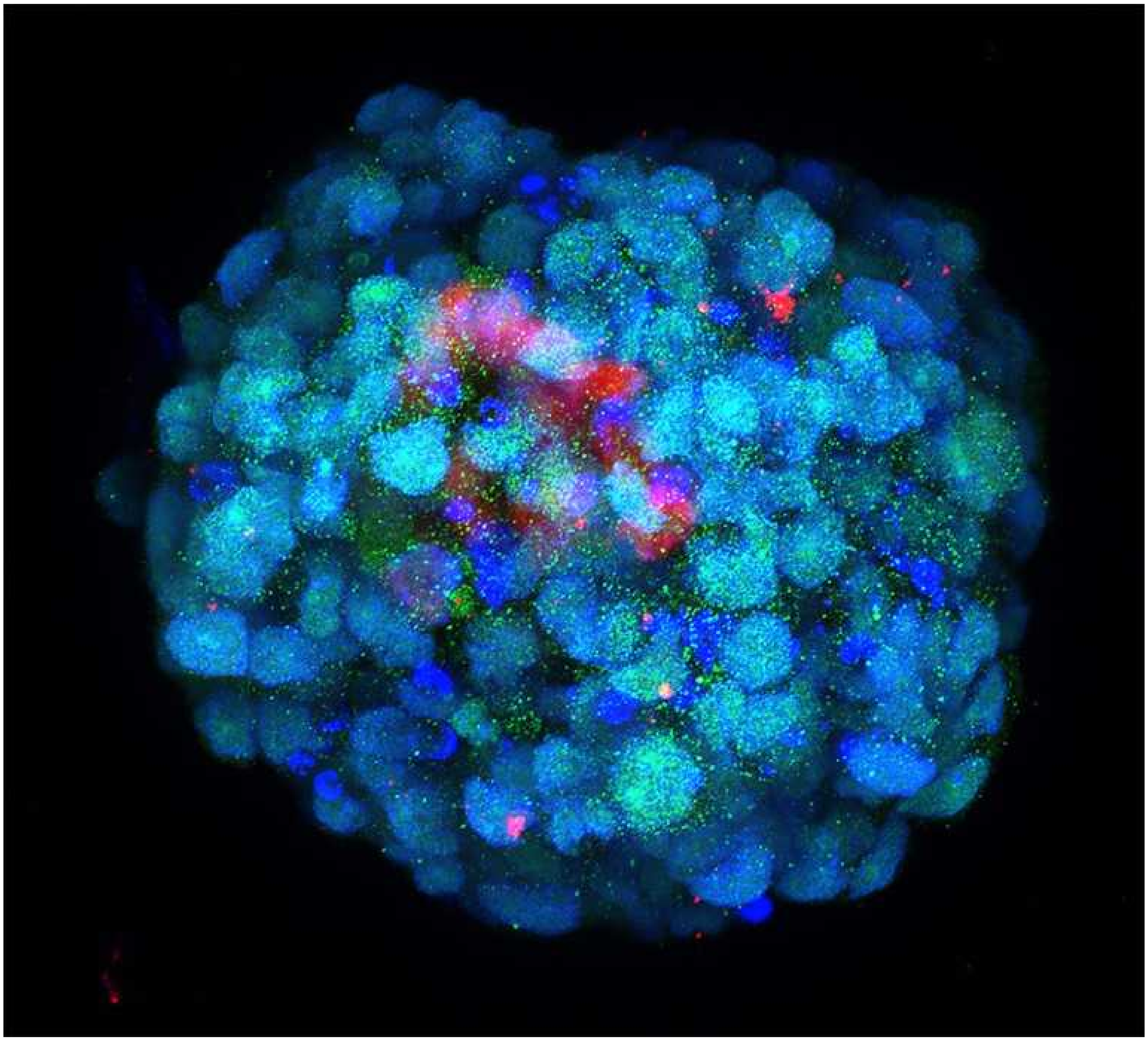
Alexandrov’s team has used SpaceM to profile hundreds of metabolites from tens of thousands of human and mouse cells, turning to standard single-cell transcriptomic methods to classify those cells into groups. Alexandrov says he is especially enthusiastic about this latter aspect and the idea of

assembling ‘metabolomic atlases’ — analogous to those developed for transcriptomics — to accelerate progress in the field. “This is definitely the frontier, and will be a big enabler,” he says.

In vitro embryo models

The journey from fertilized ovum to fully formed embryo has been mapped in detail at the cellular level for mice and humans. But the molecular machinery driving the early stages of this process remains poorly understood. Now a flurry of activity in ‘embryoid’ models is helping to fill these knowledge gaps, giving researchers a clearer view of the vital early events that can determine the success or failure of fetal development.

Some of the most sophisticated models come from the lab of Magdalena Zernicka-Goetz, a developmental biologist at the California Institute of Technology in Pasadena and the University of Cambridge, UK. In 2022, she and her team demonstrated that they could generate implantation-stage mouse embryos entirely from embryonic stem (ES) cells[12,13].



An embryoid made using cells engineered to resemble the eight-cell stage of an embryo.Credit: M.A Mazid *et al./Nature*

Like all pluripotent stem cells, ES cells can form any cell or tissue type — but they require close interaction with two types of extra-embryonic cell to complete normal embryonic development. The Zernicka-Goetz team learnt how to coax ES cells into forming these extra-embryonic cells, and showed that these could be co-cultured with ES cells to yield embryo models that mature to stages that were previously unattainable *in vitro*. "It's as faithful as you can imagine an embryonic model," says Zernicka-Goetz. "It develops a head and heart — and it's beating." Her team was able to use this model to reveal how alterations in individual genes can derail normal embryonic development[12].

At the Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, stem-cell biologist Miguel Esteban and colleagues are taking a different tack: reprogramming human stem cells to model the earliest stages of development.

“We started with the idea that actually it might even be possible to make zygotes,” Esteban says. The team didn’t quite achieve that, but they did identify a culture strategy that pushed these stem cells back to something resembling eight-cell human embryos[14]. This is a crucial developmental milestone, associated with a massive shift in gene expression that ultimately gives rise to distinct embryonic and extra-embryonic cell lineages.

Although imperfect, Esteban’s model exhibits key features of cells in natural eight-cell embryos, and has highlighted important differences between how human and mouse embryos initiate the transition to the eight-cell stage. “We saw that a transcription factor that is not even expressed in the mouse regulates the whole conversion,” says Esteban.

Collectively, these models can help researchers to map how just a few cells give rise to the staggering complexity of the vertebrate body.

Research on human embryos is restricted beyond 14 days of development in many countries, but there’s plenty that researchers can do within those constraints. Non-human primate models offer one possible alternative, Esteban says, and Zernicka-Goetz says that her mouse-embryo strategy can also generate human embryos that develop as far as day 12. “We still have lots of questions to ask within that stage that we are comfortable studying,” she says.

Nature **613**, 794-797 (2023)

doi: https://doi.org/10.1038/d41586-023-00178-y

References

1. Swaminathan, J. *et al. Nature Biotechnol.* **36**, 1076–1082 (2018).

2. Reed, B. D. *et al. Science* **378**, 186–192 (2022).

3. Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A. & Dekker, C. *Science* **374**, 1509–1513 (2021).

4. Heinrich, L. *et al. Nature* **599**, 141–146 (2021).

5. Xu, C. S. *et al. Nature* **599**, 147–151 (2021).

6. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. *et al. Science* **368**, 290–296 (2020).

7. Ciciani, M. *et al. Nature Commun.* **13**, 6474 (2022).

8. Kuitems, M. *et al. Nature* **601**, 388–391 (2022).

9. Miyake, F., Nagaya, K., Masuda, K. & Nakamura, T. *Nature* **486**, 240–242 (2012).

10. Brehm, N. *et al. Nature Commun.* **13**, 1196 (2022).

11. Rappez, L. *et al. Nature Methods* **18**, 799–805 (2021).

12. Amadei, G. *et al. Nature* **610**, 143–153 (2022).

13. Lau, K. Y. C. *et al. Cell Stem Cell* **29**, 1445–1458 (2022).

14. Mazid, M. A. *et al. Nature* **605**, 315–324 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00178-y

- WHERE I WORK
- 23 January 2023

Why I’m a garbage collector for the world’s oceans

Boyan Slat founded and operates a business that seeks to reduce ocean plastic pollution by 90%.

- Linda Nordling



Boyan Slat is the chief executive and founder of The Ocean Cleanup in Rotterdam, the Netherlands. Credit: The Ocean Cleanup

I’m the founder and chief executive of The Ocean Cleanup, a non-profit organization based in Rotterdam, the Netherlands, that aims to remove plastic pollution from the oceans. My team and I do this in two main ways: by dragging nets across areas where plastic accumulates, such as the Great Pacific Garbage Patch, a collection of debris in the North Pacific Ocean; and by intercepting waste in rivers before it reaches the sea. We want to reduce ocean plastic pollution by 90% by 2040.

This photo was taken in Guatemala near the mouth of the Motagua River, downstream of Guatemala City and a huge landfill. Every year, around 20,000 tonnes of plastic flow down the river into the Caribbean Sea. In mid-2022, we put an eight-metre-high avalanche fence across the river to catch the plastic. Soon after, heavy rains washed about one million kilogrammes of waste down the river in 20 minutes, causing the foundations of our fence to fail. In the picture, I’m assessing the aftermath of that event. We are now engineering an improved version of the barrier, which we hope will be in place before the next rainy season.



The business of science

There’s much we don’t understand about plastic pollution in the oceans. When we started, we knew about the Great Pacific Garbage Patch, but we didn’t know how big it was or how much plastic enters the oceans every year. Plastic can take decades to break down — in the Pacific garbage patch,

much of the rubbish is from the 1990s, and we have even found items from the 1960s.

The Ocean Cleanup has more than 100 staff members, including scientists. We try to be a haven for researchers who want to do applied research and make a difference. So far, we've removed around two million kilogrammes of rubbish from rivers and the ocean, including about 200,000 kilogrammes from the Pacific garbage patch. That's only a tiny amount — about 0.2% of what we estimate is in the garbage patch. But we hope to dramatically increase the amount of plastic removed.

Nature **613**, 800 (2023)

doi: https://doi.org/10.1038/d41586-023-00179-x

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00179-x

Research

- **Attosecond field emission** [25 Jan 2023]
  • By using sub-cycle light transients to induce optical field emission of electron pulses from tungsten nanotips and a weak replica of the same transient, investigation of attosecond emission dynamics in real time is demonstrated.
- **A wearable cardiac ultrasound imager** [25 Jan 2023]
  • Innovations in device design, material fabrication and deep learning are described, leading to a wearable ultrasound transducer capable of dynamic cardiac imaging in various environments and under different conditions.
- **Social trauma engages lateral septum circuitry to occlude social reward** [30 Nov 2022]
  • The authors show that, in a chronic social defeat stress rodent model, a subset of male and female mice avoided social interaction with non-aggressive, same-sex juvenile mice and did not develop context-dependent social reward following these encounters.
- **Discriminatory attitudes against unvaccinated people during the pandemic** [08 Dec 2022]
  • Vaccinated people express discriminatory attitudes towards unvaccinated individuals across cultures.
- **Undecaprenyl phosphate translocases confer conditional microbial fitness** [30 Nov 2022]
  • Members of the DUF368-containing and DedA transmembrane protein families have conditional roles in undecaprenyl phosphate translocation in Gram-negative and Gram-positive bacteria and may have a widely conserved function in the biogenesis of microbial cell surface glycopolymers.
- **Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination** [06 Dec 2022]
  • Pre-existing high-affinity antibodies alter germinal centre and memory B cell selection by lowering the activation threshold for B cells and through direct masking of their cognate epitopes, thereby permitting a diverse set of abundant lower-affinity clones targeting alternate epitopes to participate in the immune response.
- **γδ T cells are effectors of immunotherapy in cancers with HLA class I defects** [11 Jan 2023]
  • γδ T cells contribute to the response to immune checkpoint blockade treatment in patients with HLA-class-I-negative DNA mismatch repair-deficient colon cancers. .

- **An atlas of substrate specificities for the human serine/threonine kinome** [11 Jan 2023]
  • Analysis of the kinase activity of 300 protein Ser/Thr kinases reveals that the substrate specificity of the kinome is substantially more diverse than expected and is driven extensively by negative selectivity
- **Structures of the holo CRISPR RNA-guided transposon integration complex** [28 Nov 2022]
  • Structural studies of the CRISPR-associated transposon comprising Cas12k, TnsC, TnsB and TniQ from Scytonema hofmannii using cryo-electron microscopy reveal insights into the architecture and mechanism of RNA-guided DNA transposition.

- Article
- Open Access
- Published: 25 January 2023

Attosecond field emission

- H. Y. Kim,
- M. Garg,
- S. Mandal,
- L. Seiffert,
- T. Fennel &
- …
- E. Goulielmakis

Show authors

Nature **volume 613**, pages 662–666 (2023)

- 3297 Accesses

- 131 Altmetric

- Metrics details

Abstract

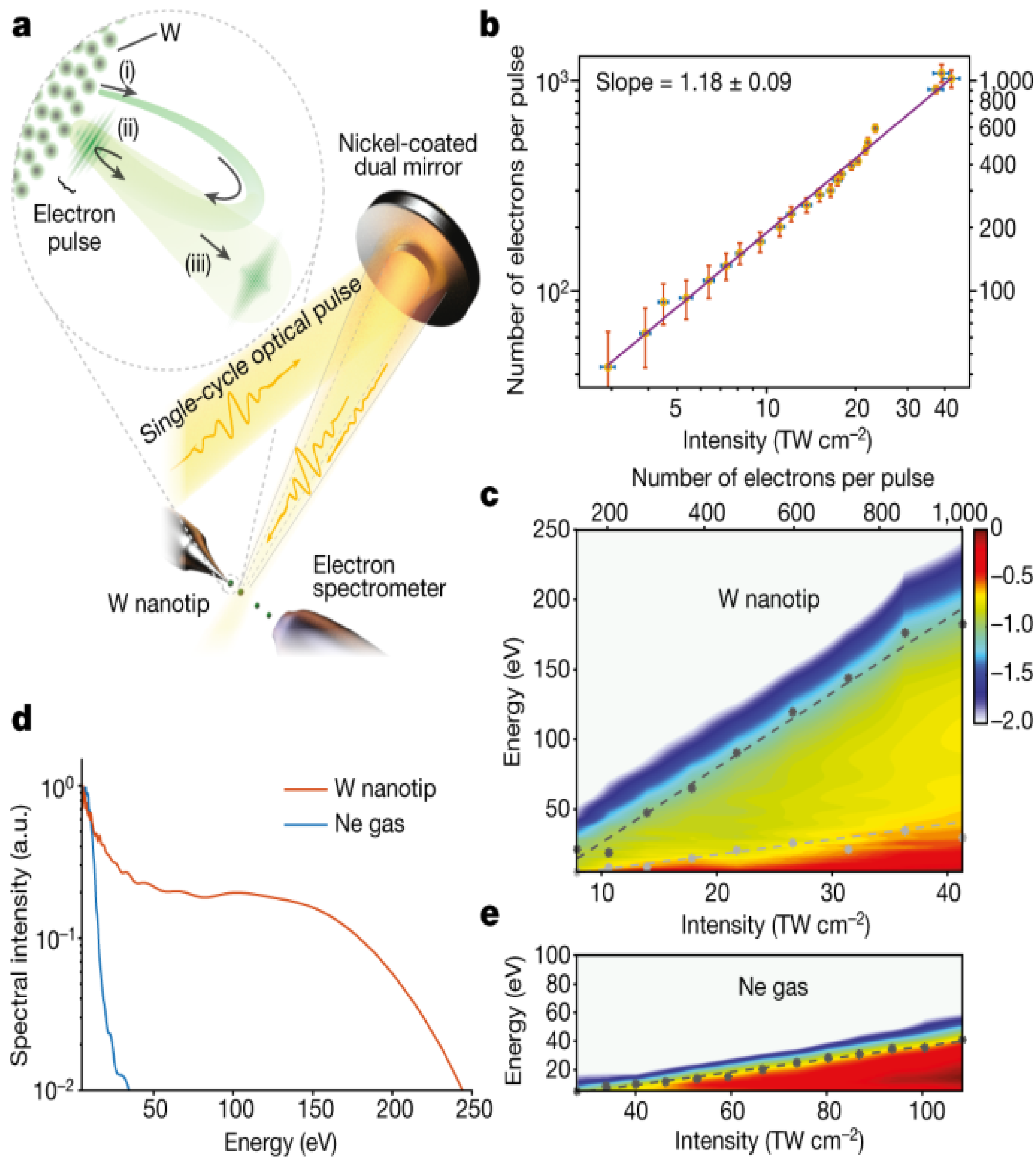
Field emission of electrons underlies great advances in science and technology, ranging from signal processing at ever higher frequencies[1] to imaging of the atomic-scale structure of matter[2] with picometre resolution. The advancing of electron microscopy techniques to enable the complete visualization of matter on the native spatial (picometre) and temporal (attosecond) scales of electron dynamics calls for techniques that can confine and examine the field emission on sub-femtosecond time intervals. Intense laser pulses have paved the way to this end[3,4] by demonstrating

femtosecond confinement[5,6] and sub-optical cycle control[7,8] of the optical field emission[9] from nanostructured metals. Yet the measurement of attosecond electron pulses has remained elusive. We used intense, sub-cycle light transients to induce optical field emission of electron pulses from tungsten nanotips and a weak replica of the same transient to directly investigate the emission dynamics in real time. Access to the temporal properties of the electron pulses rescattering off the tip surface, including the duration $\tau = (53\ \text{as} \pm 5\ \text{as})$ and chirp, and the direct exploration of nanoscale near fields open new prospects for research and applications at the interface of attosecond physics and nano-optics.

Main

The interaction of atoms and molecules with intense laser fields gives rise to attosecond electron pulses[10] that can study the structure and dynamics of these systems on recollision with their parent ion[11]. Attosecond techniques[12] can now gain access to the temporal profile of the recolliding electron pulses and concomitant structural dynamics[13,14] in their parent ions by measuring the transient properties of high harmonics[15] emitted during the interaction. Studies of the interaction of intense laser fields with nanostructured metals over the past two decades have suggested that the semiclassical concepts[3,4,5,6,7,8,9,16,17] developed earlier to describe electron dynamics in atoms can afford a central role in the understanding of the optical field electron emission. In analogy to atoms, electrons set free from the apex of a nanotip at the field crest of an intense laser pulse should also form ultrashort electron pulses (Fig. 1a, inset), which—on recollision with the tip surface about three-quarters of the laser period ($T \approx 2$ fs) later—could examine both dynamics and structure. Owing to the ultrashort time interval between emission and recollision events, and in contrast to other emerging electron-pulse technologies[18,19,20,21], the electron-pulse wave packet shall undergo a negligible temporal spread, allowing its confinement to sub-cycle timescales.

Fig. 1: Optical field emission by intense, sub-cycle optical transients.



a, Simplified schematic of the experimental setup. A sub-cycle pulse (orange curve) is spatially separated and focused by a dual concave Ni mirror module. A time delay between the pulses reflected by the inner and outer mirrors is introduced by a piezo stage. Tungsten nanotips (apex radius around 35 nm) or a gas jet of neon atoms can be positioned in the laser focus. Emitted electron spectra are recorded by a time-of-flight spectrometer (acceptance angle about 6°) placed roughly 3 mm downstream of the electron source and aligned along the laser polarization axis. The inset

shows that electrons marked by the green shaded curve are (i) set free and accelerated by the intense laser field to form an electron pulse, which, on recollision with the nanotip surface, (ii) can investigate both dynamics as well as structure. On backscattering off the tip surface, (iii) the electron pulse is further accelerated by the laser to escape the interaction area. **b**, Total electron yield per pulse as a function of increasing peak intensity of the driving laser pulses (yellow points) and its linear fitting (purple line) on a logarithmic scale. **c**, Electron spectra from the tungsten nanotip (false colour in logarithmic scale) versus peak intensity. Stars and dots denote the cutoff energies. Black and grey dashed lines show the linear fitting of the cutoff energy versus incident peak intensity of the laser pulse. **d**, Optical emission electron spectra from the tungsten nanotip (red curve) and neon atoms (blue curve) for nearly identical peak intensity (about 40 TW cm^{-2}). **e**, Same as in **c** but for neon atoms.

Source data

However the real-time tracking of electron pulses generated in the optical field emission has remained challenging. Although ordinary attosecond streaking techniques can be used to map the temporal structure of the extreme ultraviolet (EUV) electron emission from metal surfaces[22] and nanotips[23], they cannot directly investigate electron pulses emerging in the optical field emission. Furthermore, the absence of high harmonic emission from laser-driven nanotips constrains the applicability of in situ attosecond techniques[24,25,26,27] for examining the structure of the electron pulses in these systems.

Harnessing electron pulses emerging in the optical field emission for realizing new spectroscopies that combine attosecond temporal and nanometre spatial resolution calls for essential advancements on both their generation and their measurement methodologies. On the generation side, the driving laser pulses should be both short and intense so as to confine the tunnelling of electrons into a sub-femtosecond window (<1 fs), as well as to impart the recolliding electron pulses with a de Broglie wavelength (<2.74 Å, >20 eV) that allows atomic-scale investigation of the parent surface[28]. On the measurement side, in situ attosecond metrologies[24,25,26,27] shall be extended to incorporate temporal gating of the optical field emission without

relying on the concomitant high harmonic radiation. Measurements of this kind have so far permitted access to the driving-field waveform of light waves by tracking the spectrally integrated currents induced in the bulk of solids[29,30] or the cutoff energy variation of rescattered electrons in atoms[31] and nanotips[32], but a direct time-resolved measurement of attosecond electron pulses in the optical field emission has remained beyond reach.

Guided by the above provisions, we studied the field emission in tungsten nanotips (work function $\phi \approx 4.5$–5.0 eV) using intense (around 10^{13} W cm^{-2}), sub-cycle (about 1.9 fs) optical transients (centroid energy around 1.8 eV) generated in a light-field synthesizer[33,34]. The experiments were performed in a multifunctional experimental setup (see Methods and Extended Data Fig. 1) that combines photoemission spectroscopy of atoms and solids, optical pump–probe measurement methodologies and EUV attosecond streaking[35] for the sampling of the driving-field waveforms.

We began our experiments by analysing the nonlinearity of the electron yield to the driving-field intensity. A plate detector (see Extended Data Fig. 2) recorded the total number of emitted electrons from the electrically grounded nanotip (yellow points, Fig. 1b) for a range of peak intensities of the impinging laser transient. For the highest peak intensity of around 42 TW cm^{-2}, we recorded up to about 1,000 electrons per pulse. Evaluation of the slope of the total electron yield versus peak intensity (purple line in Fig. 1b) showed an emission nonlinearity of about 1.18 ± 0.09, which is well below the multiphoton threshold $\phi/\hbar\omega_L \approx 3$ and supports the notion that field-driven electron tunnelling dominates the ionization of tungsten over the entire range of the studied intensities.

Next we conducted a spectral-domain study of the emission under precisely characterized driving fields and interrogated compatibility of our findings with the predictions of semiclassical, single-electron models. Given the multielectron nature of the emission (Fig. 1b), this step is essential for applying semiclassical single-electron concepts later in this work to scrutinize the temporal structure of the electron emission. As space–charge interactions can substantially distort the emission spectra, in comparison with semiclassical predictions, a detailed interrogation of the cutoff energies of the emission can shed light onto potential multielectron contributions.

Electron spectra recorded as a function of the peak intensity of the driving pulse (Fig. 1c) showed the formation of two well-discernible cutoffs (black points, grey points and false-colour plot in Fig. 1c, respectively), whose energies scale linearly with peak intensity. Evaluation of the corresponding slopes for high and low cutoff energies (E_c) by linear fitting (black and grey dashed lines in Fig. 1c) of the experimental data yielded ${s}_{\text{W},\text{high}}^{(\exp)}={\rm{d}}{E}_{{\rm{c}}}/{\rm{d}}{U}_{{\rm{p}}}=118\pm 5.1$ and ${s}_{\text{W},\text{low}}^{(\exp)}={\rm{d}}{E}_{{\rm{c}}}/{\rm{d}}{U}_{{\rm{p}}}=24.1\pm 1.35$, respectively. Here U_p is the ponderomotive energy. By taking the ratio between high and low cutoff energy slopes (about 4.91 ± 0.35), we find that it is compatible with that anticipated for backscattered and direct electrons ($10U_p/2U_p \approx 5$) in the single-electron approximation[36].

An experimental assessment of the near-field enhancement in the vicinity of the nanotip and its comparison with the theoretical predictions could further attest to the validity of single-electron semiclassical pictures to describe emission under our experimental conditions. To this end, we compared the emission spectra of electrons from the nanotip with those in a low-density neon gas. Figure 1d contrasts electron spectra recorded from W nanotip (red curve) and neon atoms (blue curve) under identical field waveforms. Figure 1e shows electron spectra emanating from Ne over a broad range of driving-pulse intensities. A linear fitting (Fig. 1e, dashed line) of the cutoff energy (Fig. 1e, points) of the neon spectra versus intensity yielded a slope ${s}_{{\rm{Ne}}}^{(\exp)}=9.88\pm 0.34$ that agrees well with the semiclassical predictions under our driving waveforms ($s_{Ne} = 10.8$) (see Methods). The field-enhancement factor in the vicinity of the tungsten nanotip, which is now evaluated as $f=\sqrt{\frac{{s}_{{\rm{W}},{\rm{high}}}^{(\exp)}}{{s}_{{\rm{Ne}}}^{(\exp)}}}=3.46\pm 0.10$, shows a fair agreement with the theoretical prediction, $f_{th} \approx 3.8$ (Methods). This comparison further attests to the compatibility of the emission processes from the nanotip with single-electron, semiclassical concepts and suggests that multielectron charge interactions remain negligible.

Detailed semiclassical simulations based on the experimentally derived quantities (see Methods) further support the above perspective. The

simulations accurately reproduce the experimental spectra (Extended Data Fig. 4) and the association of high and low cutoff energies of the emission to backscattered and direct electrons, respectively. The inclusion of multielectron interactions (Extended Data Fig. 6) showed the presence of weak-only effects of space–charge distortions on the emitted spectra and the spectral phase of the electron emission further corroborating the appropriateness of single-electron models to describe the emission process.

The absence of discernible cutoffs at intermediate energies (60–120 eV) in both experiments (Fig. 1c) and simulations (Extended Data Figs. 4 and 6) supports the notion that, under sub-cycle driving, the high-energy part of the emitted spectrum is associated with the recollision of a single electron pulse at the tip surface. A weak, low-energy backscattered emission indicated by the simulations (Extended Data Fig. 4a) is not directly resolved in the experiments (Fig. 1c). Yet the presence of such emission channels will become apparent later in this work by time resolving the optical field emission.

Homochromatic attosecond streaking

As vacuum is dispersive to electron pulses, a temporal characterization has a concrete meaning at a specific point in space. Because the generated electron pulses scrutinize the 'sample' during the recollision with the parent surface as to their temporal structure at this point, it is mostly relevant for harnessing the power of these pulses in time-resolved applications.

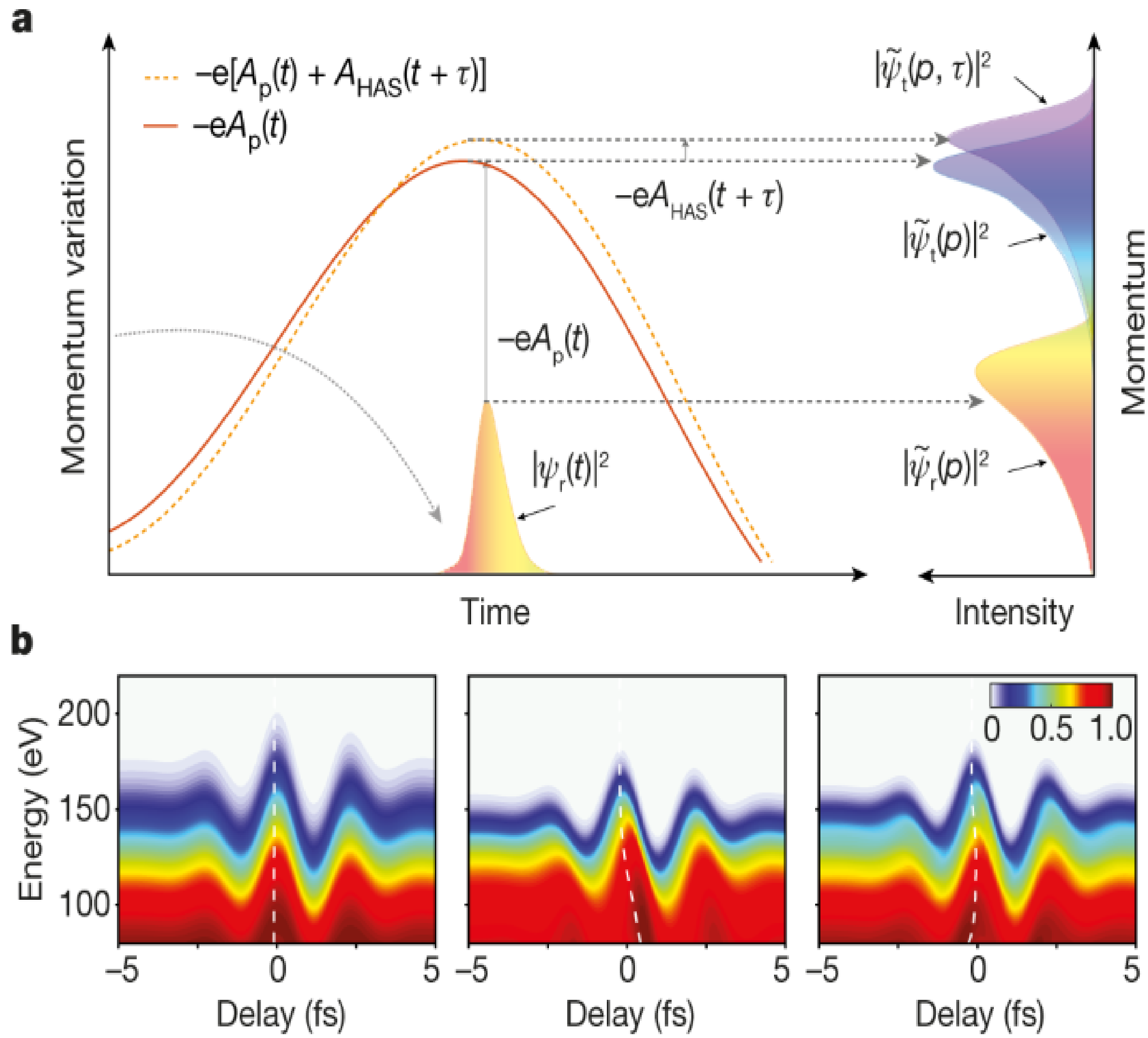
To understand how, we revisit the process of strong field recollision of an electron wave packet under an intense optical waveform (Fig. 2). Set free around the peak of a laser field crest, an electron pulse will recollide with the tip surface at an instance t_{r} (Fig. 2a) with an energy of about $3U_{\mathrm{p}}$ (refs. [11,37]). An attosecond measurement of the electron pulse entails access into its waveform $\psi_{\mathrm{r}}(t)$ at the surface of the nanotip or equivalently, into its associated complex spectral amplitude ${\widetilde{\psi }}_{{\rm{r}}}(p)$, in which p is the recollision momentum of the electron. However, as ${\widetilde{\psi }}_{{\rm{r}}}(p)$ is not directly accessible in measurements, it is important to link it to other measurable quantities. Following backscattering off the tip surface, the wave packet acquires

further phase from both its interaction with the driving field (Volkov phase) as well owing to its free-space propagation. If we define an auxiliary terminal wave packet ${\widetilde{\psi }}_{{\rm{t}}}(p)$, the spectral intensity $I(p)={| {\widetilde{\psi }}_{{\rm{t}}}(p)| }^{2}$ (refs. [38,39,40]) (as marked in Fig. 2a) can be expressed as (in atomic units):

$$I(p)={|{\widetilde{\psi }}_{t}(p)|}^{2}={\left|{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\psi }_{r}({t}_{r}){\rm{\exp }}\left[i\frac{{p}^{2}}{2}{t}_{r}\right]{\rm{\exp }}[-{iS}(p,{t}_{r}{\rm{;}}{A}_{p}(t))]{\rm{d}}{t}_{r}\right|}^{2}$$
(1)

in which $S(p,{t}_{{\rm{r}}}\,;{A}_{{\rm{p}}}(t))={\int }_{{t}_{{\rm{r}}}}^{{\rm{\infty }}}\left[\frac{1}{2}\right.{[p+{A}_{{\rm{p}}}(t)]}^{2}-\left.\frac{1}{2}{p}^{2}\right]\,{\rm{d}}t$ denotes the Volkov phase imparted to the electron wave packet by the vector potential $A_p(t)$ of the intense driving pulse (hereafter referred to as pump, see red curve in Fig. 2a) after rescattering at time t_r. Hence reconstruction of the recolliding wave packet $\psi_r(t)$ should be possible if, other than $I(p)$, which is a directly measurable quantity (that is, the spectrum of the electron emission), the phase of ${\widetilde{\psi }}_{{\rm{t}}}(p)$ as well as the waveform $A_p(t)$ are accessed.

Fig. 2: Homochromatic attosecond streaking.



a, An electron wave packet $\psi_{\rm r}(t)$, sampled on recollision at the nanotip surface, is represented by its momentum distribution $\widetilde{\psi}_{\rm r}(p)$ and corresponding spectral intensity $\left|\widetilde{\psi}_{\rm r}(p)\right|^{2}$. Following backscattering, further acceleration of the wave packet by the vector potential of the pump field $A_{\rm p}(t)$ (red curve) gives rise to a terminal wave packet $\widetilde{\psi}_{\rm t}(p)$ with corresponding spectral intensity $I(p)=\left|\widetilde{\psi}_{\rm t}(p)\right|^{2}$. The coherent superposition (dashed yellow curve) of the pump field with a weak, time-delayed (τ) replica (the gate) results in a momentum shift, $-eA_{\rm HAS}(t+\tau)$ of the final momentum distribution $I(p,\tau)=\left|\widetilde{\psi}_{\rm t}(p,\tau)\right|^{2}$. **b**, Simulated delay-dependent homochromatic attosecond streaking (HAS) spectrograms for electron pulses without chirp

(left panel), positive chirp of 3.5×10^3 as^2 (middle panel) and third-order chirp of 1×10^5 as^3 (right panel). The dashed white curves highlight the resulting energy-dependent shifts of the intensity modulation.

Source data

Access to the phase of ${\widetilde{\psi }}_{{\rm{t}}}(p)$ is possible by temporally gating the electron emission with a weak replica of the pump pulse (hereafter referred to as gate with vector potential $A_g(t)$) when $\eta \equiv {\left|{A}_{{\rm{g}}}(t)\right|}^{2}/{\left|{A}_{{\rm{p}}}(t)\right|}^{2}\ll 1$ (refs. [27,31]). In this limit, the pump pulse is solely responsible for releasing the electron wave packet, whereas the gate pulse primarily alters its phase. This is manifested by the shift and reshaping of the terminal photoelectron spectra at the end of the driving pulse (Fig. 2a). If the delay τ between pump and gate pulses is varied (see Methods), the terminal spectral distribution of the released electron, $I(p,\tau)={\left|{\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\right|}^{2}$ can be approximated as:

$$I(p,\tau)\propto {\left|{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\psi }_{t}\left({t}_{r}\right){\rm{\exp }}\left[i\frac{{p}^{2}}{2}{t}_{r}\right]{\rm{\exp }}\left[-{iS}\left(p,{t}_{r}{\rm{;}}{A}_{{HAS}}\left(t\right)\right)\right]{\rm{d}}{t}_{r}\right|}^{2}$$
(2)

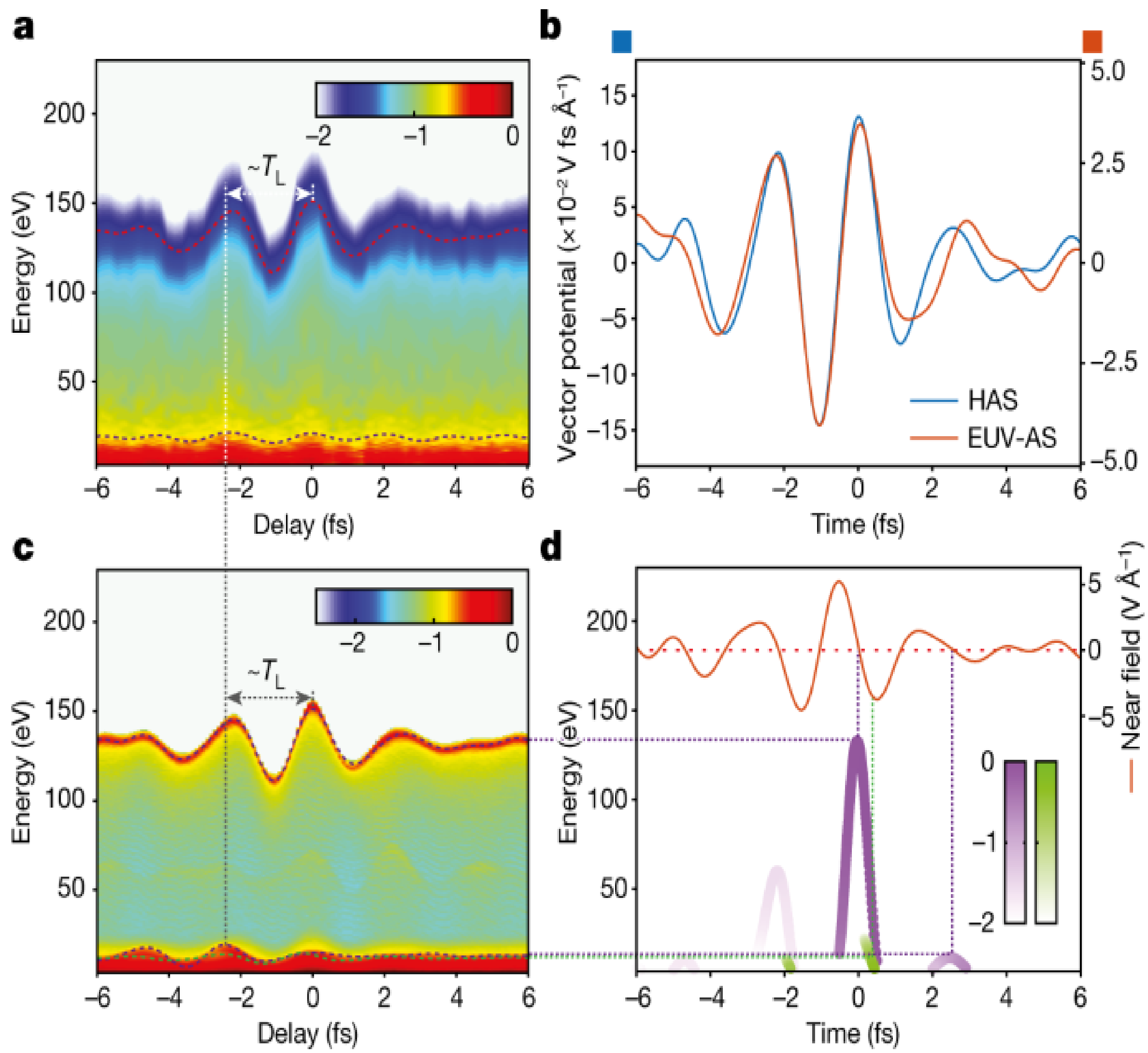
in which $\psi_t(t)$ is the inverse Fourier transform of ${\widetilde{\psi }}_{{\rm{t}}}(p)$ and $A_{HAS}(t+\tau)$ represents an effective vector potential, which is explicitly related to the incident vector potential $A_g(t)$ of the gate pulse as shown in Methods and which accounts for the momentum an electron accumulates from the instance of its birth to the detection. Equation (2) implies: (i) a variation of τ permits the composition of a streaking-like spectrogram whose reconstruction can allow retrieving the phase of ${\widetilde{\psi }}_{{\rm{t}}}(p)$ and (ii) the momentum variation of the electron distribution follows $A_{HAS}(t)$. Whereas implication of (ii) (see Methods) allows sampling of the waveform of a light pulse[31], (i) is essential for mapping the dynamics of the field emission. To distinguish from conventional attosecond streaking, we refer to this approach as

homochromatic attosecond streaking (HAS), highlighting that the carrier frequency of the pump and gate fields is identical.

Shown in Fig. 2b are simulated HAS spectrograms under conditions pertinent to the experiments presented here. Notably, and in close analogy to ordinary attosecond streaking[41,42], different types of chirp of the recolliding electron pulse $\psi_r(t)$ yield distinct visual signatures in the spectrogram, manifested as shifts of the spectral intensity modulation versus delay and energy, as highlighted by the dashed white curves in Fig. 2b.

In our experiments, we derived pump and gate pulses by the spatiotemporal division of the sub-cycle optical transients using a dual-mirror module as shown in Fig. 1a. Figure 3a plots a HAS spectrogram recorded by our apparatus. A ratio $\eta \approx 6.3 \times 10^{-3}$ warrants that the gate pulse is sufficiently weak to serve as a phase gate (see Methods), whereas a remaining weak amplitude modulation (5–10%) of the spectrogram versus delay was useful for evaluating the timing between pump and gate pulses, and therewith to clock the recollision instance with respect to the waveform of the driving pump pulse.

Fig. 3: Homochromatic attosecond streaking of the optical field emission.



a, Homochromatic attosecond streaking (HAS) spectrogram of the optical field emission from a tungsten nanotip consisting of 60 individual electron spectra recorded as a function of the delay (step size of 200 as) between an intense ($I_p \approx 25.4$ TW cm^{-2}) single-cycle pulse and a 6.3×10^{-3} times weaker gate pulse. **b**, Vector potential $A_g(t)$ evaluated by the photoelectron energy cutoff variation in the HAS spectrogram in panel **a** as a function of the delay (blue) and from an ordinary streaking spectrogram (red) (see Extended Data Fig. 9) in absolute units. **c**, Simulated HAS spectrogram using the experimentally recorded field waveforms of pump and gate pulses. The dashed vertical lines in **a** and **c** denote different rescattering events. **d**, Terminal energies of direct (green) and backscattered electron emissions (purple) associated (dashed horizontal lines) with corresponding cutoff energies in the spectrogram (**c**), as well as release times (dashed vertical

lines) near-field pump waveform (red curve). EUV-AS, extreme ultraviolet attosecond streaking.

Source data

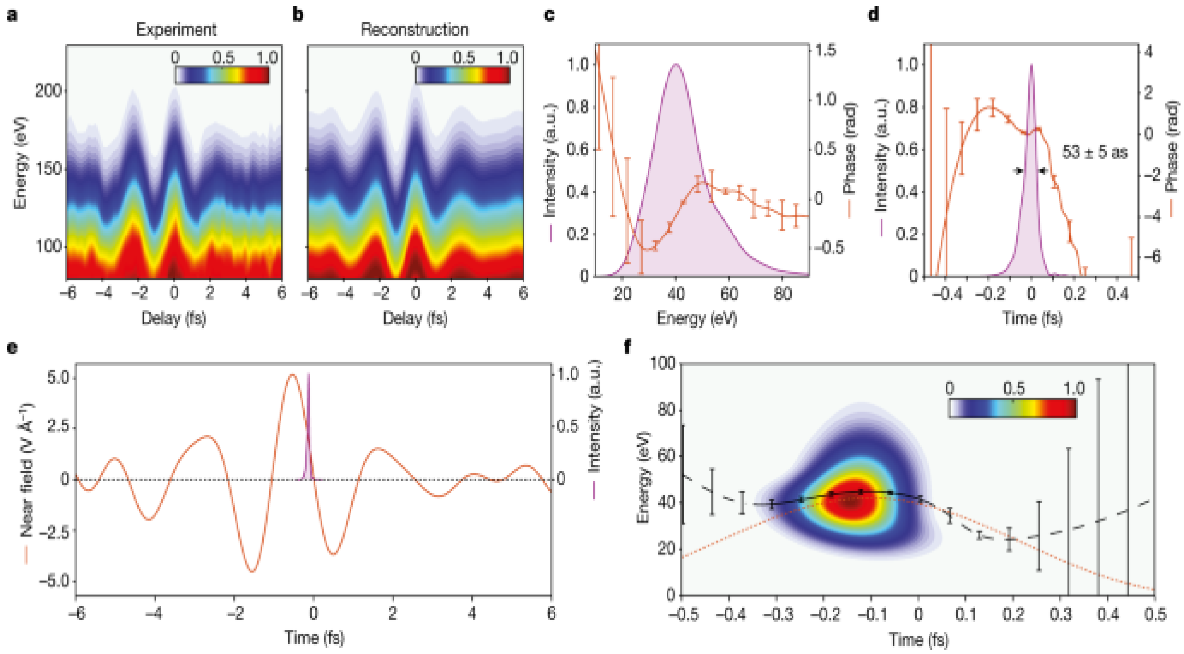
The evaluated $A_g(t)$ (blue curve in Fig. 3b) by HAS exhibits an excellent waveform matching that derived by EUV attosecond streaking (red curve in Fig. 3b), as attested by the degree of similarity[43] (around 0.95). In this case, the HAS-evaluated vector potential represents the near field of the gate pulse in the vicinity of the tip. Moreover, the ratio between the absolute amplitudes of the vector potentials evaluated by the two methods provides a direct measurement of the field-enhancement factor $f = 3.73 \pm 0.25$, which is in close agreement with the result of the methodology highlighted in Fig. 1c (conducted with a different nanotip) and the theoretical estimations.

How time-domain measurement of electron emission in nanostructured materials benefits the intuitive understanding of the process and its real-time clocking can be best appreciated by a close inspection of the oscillating phases of high and low cutoff energies in Fig. 3a. For instance, an apparent delay of the maxima of the corresponding oscillations (highlighted by red and blue dashed curves, respectively, in Fig. 3a) versus delay indicates a retardation in the electron emission at lower energies by approximately a laser cycle ($T_L \approx 2.3$ fs). This feature, also well reproduced in our simulation (Fig. 3c), shows that the low-energy emission consists of a mixture of direct electrons emerging within the main half-cycle of the driving field (green dots in Fig. 3d) and low-energy backscattered electrons generated approximately a cycle of the driving field later (purple dots in Fig. 3d). Moreover, the uniform amplitude and energy modulation of the photoelectron spectrum (Fig. 3a) over a broad range of energies (50–150 eV) offers compelling evidence that emission is confined to a single electron pulse generated within a field half-cycle.

For the reconstruction of the electron wave packet, we are primarily interested in the properties of the recolliding electron pulses at energies typically higher than 20 eV, that is, where this pulse could serve high-resolution/atomic-scale, spatial analysis of surfaces[28]. Considering the ratio of the terminal energy of electrons (about $10U_p$) and that at the recollision

instance (about $3.17U_p$), we conclude that the relevant information is encoded at the high-energy end (>80 eV) of the spectrogram of Fig. 3a, as isolated in Fig. 4a. Figure 4b shows the numerical reconstruction of data in Fig. 4a based on equations (1) and (2), the retrieved field parameters $A_p(t)$ and $A_g(t)$ (Fig. 3b), the absolute time delay τ (Extended Data Fig. 8) and the numerical algorithm detailed in Methods.

Fig. 4: Measurement of attosecond electron pulses in the optical field emission.



a,**b**, Experimentally recorded (**a**) and reconstructed (**b**) HAS spectrograms from a W nanotip. **c**, Retrieved backscattered electron-pulse spectrum (magenta fill) and its spectral phase (red curve). **d**, Intensity profile of the electron pulse (magenta fill) and its temporal phase (red line). **e**, Near field (red curve) and its timing with respect to the attosecond electron pulse (magenta fill). **f**, Time–frequency analysis of the attosecond electron pulse (false colour is intensity in arbitrary units) and retrieved release times (black curve). The dashed red curve denotes release times calculated semiclassically for the near-field light waveform in **e**. Error bars represent standard deviations of the corresponding values indicated.

Source data

The retrieved electron-pulse profiles in spectral and time domains are shown in Fig. 4c,d, respectively. The spectrum extends over the range 20–80 eV (Fig. 4c, magenta-shaded curve) and is temporally confined to about 53 ± 5 as, as measured at the full width at half maximum of its intensity profile (Fig. 4d, magenta-shaded curve). Figure 4e compares the retrieved electron pulse and the near-field profiles, and shows that the electron wave packet revisits the surface of the nanotip (magenta-shaded curve) at times close to the zero transition of the driving field (red curve). This observation is compatible with the semiclassical understanding of strong-field rescattering in atoms[11,36].

A closer inspection of the retrieved spectral and temporal phases (red curves in Fig. 4c,d) also shows a negligible temporal spread of the electron pulse compared with its Fourier-limited duration (about 50 as). The compatibility of this finding with the semiclassical model can be best appreciated by comparing the group delay of the electron release (solid black line in Fig. 4f) as evaluated by the time–frequency analysis (Fig. 4f, false colour) of the attosecond electron-pulse waveform in Fig. 4d with the semiclassical release times (dashed red curve in Fig. 4f) calculated using the near-field waveform of the driving pulse (red curve in Fig. 4e). The comparison highlights that, as anticipated by the semiclassical recollision model, electrons of energy close to the cutoff, that is, where short and long trajectories merge, carry a minor temporal chirp. This feature, common in near-cutoff high harmonic emission in atoms[24,41], further strengthens the link between strong-field optics in atoms and nano-optics systems.

Conclusion

Direct measurement of attosecond electron pulses in the optical field emission extends the repertoire of ultrafast-science tools. The technique presented here could offer a broad range of capabilities for tracking ultrafast collective or correlated electron dynamics in various materials deposited or attached to tungsten nanotips[44,45], as well as in gas-phase systems. The recolliding electron pulses shall also enable attosecond-resolved, laser-induced electron diffraction[46] experiments at nanotips, as well as nanodiffraction ex situ experiments in specimens placed at nanometric distance from the tip apex[47]. Such possibilities could enable new routes for

exploring the structure and dynamics of condensed matter in four dimensions.

Methods

Experimental

Attosecond EUV streaking

For attosecond EUV streaking measurements (Extended Data Fig. 1a), the sub-cycle transients are focused onto the neon gas jet to generate EUV pulses by high-harmonic generation. The collinearly propagating EUV and optical pulses are spatially separated by a Zr filter, which also acts as an EUV high-pass spectral filter (>70 eV), enabling the isolation of a single attosecond pulse. The EUV and optical pulses are reflected off a dual-mirror assembly, which consists of a Mo/Si inner mirror (centred at around 85 eV) and a nickel outer mirror, respectively. Inner and outer mirrors can be delayed with nanometric resolution (Extended Data Fig. 1a). EUV and optical pulses are focused onto a second Ne gas jet. Photoelectron spectra recorded as a function of the delay between the inner and outer mirrors allow the composition of attosecond streaking spectrograms, which allow the detailed characterization of the attosecond EUV pulse and, notably for these experiments, the field waveform of the optical pulse. Details about the relevant techniques can be found in refs. [35,41,48].

HAS

HAS measurements are performed on the same setup by means of (1) automated removal of the Ne gas used to generate high harmonics and the Zr filter, (2) automated replacement of the inner mirror in the dual-mirror module by a Ni-coated one of the same focal lengths (Extended Data Fig. 1b) and (3) the streaking gas nozzle is replaced by an electrically grounded tungsten nanotip. The above setup modifications are executed in a fraction of a minute and warrant identical conditions for all relevant measurements. In the HAS configuration of the setup in Extended Data Fig. 1b, Ni-coated inner and outer mirrors spatially divide the optical pulse into pump (inner

mirror beam) and gate (outer mirror) pulses. A delay between the pulses reflected off the inner and outer mirrors, respectively, is introduced by a precision transitional stage (see inset in Extended Data Fig. 1b). The HAS measurements as well as the intensity-scaling measurements of the electron spectrum and yields were performed with driving-field intensities below a critical intensity, at which the irreversible optical damage of the nanotips occurs. This upper limit of intensity was experimentally identified by observing sudden and irreversible drops of electron cutoff energy in intensity-scaling measurements.

Measurement of the absolute electron yield in the optical field emission

For the measurement of the total electron counts per pulse generated in our setup, a thin electrode (size roughly 5 mm × 5 mm) is introduced about 2 mm above the nanotip (Extended Data Fig. 2). This configuration allows detection of released electrons over a solid angle $\Omega > \pi$ steradians. The induced voltage on the thin plate is measured by a lock-in amplifier at the reference frequency of the repetition rate of the driving laser (about 3 KHz). The electronic current is evaluated by dividing the induced voltage by the system impedance (10 MΩ). The total electron count per pulse is in turn obtained by dividing the current by the repetition rate of the laser and the electron charge.

Sample robustness and measurement stability study

Tungsten nanotips were exposed to intensities of up to $I \approx 45$ TW cm^{-2} without noticing any damage. This was verified by performing the intensity–cutoff energy study shown in Fig. 1b, of both increasing as well as decreasing intensity, and recorded identical curves. When the critical intensity reached $I \approx 45$ TW cm^{-2}, the tip is damaged and the cutoff energy irreversibly drops to much lower values without the possibility for recovery unless a new tip is installed.

To explore short-term and long-term stability of our system, we recorded electron spectra as a function of time under identical conditions and for intensities typically higher than those used in the HAS measurements. Data shown in Extended Data Fig. 3 suggest an excellent stability of cutoff energy

and electron yield, implying the structural robustness of the nanotip on the timescale of typical measurements (a few minutes) as well as over several hours.

One-dimensional, semiclassical simulations of the optical field emission

The time-dependent ionization probability from a tungsten nanotip was approximated by the Fowler–Nordheim formula as[5,17,49,50]:

$$p(E(t))=N\theta (-E(t)){\left|E(t)\right|}^{2}\,\exp \,\left[-\frac{4\sqrt{2m}{\phi }^{3/2}}{3\hbar e\left|E(t)\right|}\right],$$
(3)

in which $E(t)$ is the electric-field waveform of the driving pulses, ϕ is the work function of the metal and m, $\hbar$ and e are electron mass, reduced Planck's constant and electron charge, respectively. We calculated electron trajectories using the classical equations of motion in the single-electron limit as[7,8,11,17,36]:

$$m\frac{{{\rm{d}}}^{2}{z}_{i}}{{\rm{d}}{t}^{2}}=-e{f}_{0}E(t)$$
(4)

Here i is the index of each individual trajectory and f_0 is the field-enhancement factor. At the end of the pulse, an electron spectrum is evaluated by a spectral binning of the energies of all trajectories weighted by the ionization rate at the instances of their births.

For the experiments described here, we simulated electron spectra from tungsten ($\phi = 4.53$ eV) versus peak field intensity of the driving pulse (Extended Data Fig. 4). The driving field (red curve in Extended Data Fig. 4a) used in our simulations was measured by an EUV attosecond streaking setup[35,48] integrated in our apparatus. A field-enhancement factor of $f_0 = 3.46$ used in these simulations was derived experimentally as described in the main text. The quiver length of electrons in our intensity range (the longest length of about 1.7 nm at the highest intensity of about 41 TW cm^{-2}) is considerably shorter than the decay length of the near field (about 30 nm).

Hence the released electrons experience a nearly homogeneous near field and therefore quenching effects[17] owing to the near-field decay can be neglected.

In agreement with the data of Fig. 1c, the simulated electron spectra exhibit two well-discernible energy cutoffs (red and blue dashed lines in Extended Data Fig. 4b) associated with the backscattered (purple line in Extended Data Fig. 4a) and the direct (green line in Extended Data Fig. 4a) electrons. The slopes of high ($s_{{\rm{W}},{\rm{high}}}^{({\rm{th}})}={\rm{d}}E_{{\rm{c}}}/{\rm{d}}U_{{\rm{p}}}=130.3$) and low ($s_{{\rm{W}},{\rm{low}}}^{({\rm{th}})}=26.1$) cutoff energies agree well with those in our measurements (Fig. 1c). The theory shows further emission cutoffs at energies lower than that of the direct electrons. Because these are relatively weak, they do not leave any direct signatures in the photoelectron spectra. Yet such contributions become visible in HAS spectrograms, as discussed in Fig. 3.

FDTD simulations of the field enhancement

To theoretically estimate the near-field enhancement in the vicinity of the tungsten nanotip, we numerically solved Maxwell's equations by means of three-dimensional finite-difference time-domain (FDTD) simulations. The nanotip was modelled as shown in Extended Data Fig. 5a as a sphere with radius of 35 nm that smoothly transitions to a cone with an opening angle (single side) of 15° and considering optical properties for tungsten[51]. The simulations predict a peak field-enhancement factor of about 3.8 close to the surface at the tip apex. For comparison, the spatial distribution of the enhancement at a respective tungsten nanosphere (that is, excluding the cone) is shown in Extended Data Fig. 5b, with a slightly lower peak enhancement factor of approximately 2.7.

Three-dimensional, semiclassical trajectory simulation including charge interaction

To inspect whether charge interaction substantially affects the electron emission dynamics for the considered parameters, we performed semiclassical trajectory simulations using the Mie–mean field–Monte Carlo

(M^3C) model[52]. The latter has been used extensively for the study of strong-field ionization in dielectric nanospheres[53,54,55] and has recently also been adopted for the description of metallic nanotips[56]. In brief, we mimic the apex of the nanotip as one-half of a sphere with corresponding radius. The near field is evaluated as the combined linear polarization field owing to the incident pulse (evaluated by means of the Mie solution of Maxwell's equations) and an extra nonlinear contribution owing to charge interaction treated as a mean field in electrostatic approximation (evaluated by high-order multipole expansion). The latter includes Coulomb interactions among the emitted electrons as well as their image charges (that is, an extra sphere polarization caused by the free electrons). Photoelectron trajectories are generated by Monte Carlo sampling of ionization events at the sphere surface, at which we evaluate tunnelling probabilities within WKB approximation by integration through the barrier provided by the local near field. Trajectories are propagated in the near field by integration of classical equations of motion and accounting for electron–atom collisions through respective scattering cross-sections for electrons moving within the material. To mimic the slightly higher peak enhancement of the linear response near field at a tungsten tip (≈3.8) as compared with a sphere (≈2.7), see Extended Data Fig. 5, we rescaled the incident laser intensity by a factor of 1.4. The performed M^3C simulations predict about three times fewer emitted electrons than the experiments, which we attribute to contributions of slow electrons originating from the shank of the nanotip. This is substantiated by comparison of the total electron yields predicted for the nanotip and the half-sphere, obtained through integration of the local ionization rates over the respective surface areas and the pulse duration. However the charge densities at the pole of the sphere and the tip apex are comparable, enabling to inspect the impact of charge interactions within the simplified simulation model. Extended Data Fig. 6a compares multielectron spectra simulated with (solid curves) and without (dashed curves) accounting for charge interactions among the individual electrons and for four representative settings of the peak intensities (and corresponding electron yields) of the driving field, as indicated in the legend. The presence of charge interaction is primarily manifested by a noticeable decline of the yield of direct electrons (<50 eV), which—in accordance with earlier works[52,57]—can be attributed to a partial trapping of these electrons in the vicinity of the tip. Quasi-static electric fields generated by the trapping of low-energy electrons

in turn affect the dynamics of the recolliding electrons and give rise to an increase of the terminal electron-energy cutoff[52,57]. For the highest intensity and corresponding electron yield, this shift is approximately 8% in energy.

Although transparent manifestations of such effects cannot be discerned in our experiments, it is useful to understand potential implications of charge interactions on the characterization of attosecond electron pulses using HAS. To this end, we extended our analysis to the time domain for pulses of intensity (about 31 TW cm^{-2}) and electron yield (about 600 electrons per pulse), mimicking the experimental conditions in our HAS measurements.

Extended Data Fig. 6b shows the recollision energy distribution of the electron ensemble versus release times without (top) and with (bottom) inclusion of charge interactions. The instantaneous release energy is evaluated by taking the first momentum of the time-resolved spectra (black dashed and solid curves) and the temporal phases (blue dashed and solid curves) by the temporal integration of the instantaneous energy. Their comparison, shown in Extended Data Fig. 6b, suggests marginal differences and, thus, corresponding negligible effects on the temporal characteristics of the electron pulse at the time of recollision.

When the time-domain analysis is extended to the terminal energies of the recolliding electrons versus release time, the space–charge interactions are manifested by a uniform upshift of terminal energies by about 8% but leave the temporal phase of the terminal electron wave packet unaffected (Extended Data Fig. 6c).

To investigate how this energy shift could potentially affect the retrieval of the electron pulse at the instance of recollision, we applied the simulated phase effects on our experimental data (see the section 'HAS reconstruction methodology') and evaluated once again the spectral and temporal properties of the recolliding attosecond electron pulse. Key observations associated with space–charge interactions include a weak, uniform shift of the central energy of the recolliding electron by a few electronvolts (Extended Data Fig. 6d) and a subtle change in the temporal profile of the electron pulse (Extended Data Fig. 6e), resulting in an approximately 4-as elongation of its duration at the full width at half maximum, which is within the error of the experimental reconstruction (about 5 as).

Mathematical formulation of HAS

The key objective of HAS is to retrieve the temporal structure of an attosecond electron pulse wave packet $\psi_r(t)$ at the moment of its recollision on its parent surface. As this wave function is not directly accessible, it is necessary to link it to other quantities that are either directly measured in the experiments (such as the terminal spectral intensity $I(p)={\left|{\widetilde{\psi }}_{{\rm{t}}}(p)\right|}^{2}$ at a detector) or can be reconstructed from the experimental data.

Description of strong-field electron emission

Considering an electron released from and driven back to a surface by a strong pump field $E_p(t)$, its recolliding wave packet $\psi_r(t)$ can be linked to its terminal spectral amplitude ${\widetilde{\psi }}_{{\rm{t}}}(p)$ at the end of interaction with the driving pulse. The time-dependent recollision wave packet $\psi_r(t)$ is expressed by means of its Fourier representation ${\widetilde{\psi }}_{{\rm{r}}}(p)={\int }_{-\infty }^{\infty }{\psi }_{{\rm{r}}}({t}_{{\rm{r}}})\exp [{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}]{\rm{d}}{t}_{{\rm{r}}}$ and, following the recollision, the spectral amplitude is transformed to the terminal form[38,39,40,58] (in atomic units):

$${\widetilde{\psi }}_{{\rm{t}}}(p)\propto {\rm{i}}{\int }_{-\infty }^{\infty \,}{\psi }_{{\rm{r}}}({t}_{{\rm{r}}})\exp \,\left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\,\exp [-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{p}}}(t))]{\rm{d}}{t}_{{\rm{r}}}$$
(5)

Here S is the Volkov phase imparted to the electron wave packet only by the vector potential $A_p(t)$ of the pump pulse after recollision at an instance t_r, at which the general form of the Volkov phase accumulated from a time instance t_1 to a later instance t_2 by an electric field with vector potential $A(t)$ is expressed as[59]:

$$S(p,{t}_{2},\,{t}_{1}{\rm{;}}A(t))={\int }_{{t}_{1}}^{{t}_{2}}\left[\frac{1}{2}\right.{\left[p+A(t)\right]}^{2}-\left.\frac{1}{2}{p}^{2}\right]{\rm{d}}t$$
(6)

Note that equation (6) excludes free-space propagation, that is, it vanishes in the absence of the field, and equation (5) reflects the momentum-dependent wave function at the surface, including phases accumulated only by the pump field.

Earlier semiclassical theories of strong-field emission[39,40,58] in atoms have suggested that the recolliding wave packet $\psi_r(t)$ can be expressed by integration over ionization instance t' before recollision at time t_r and over canonical momenta p' in terms of the ionization amplitude, dictated by the dipole transition $E_p(t')d(p' + A_p(t))$, the scattering amplitude typically described as $g(p' + A_p(t_r))$ and the Volkov phase that the electron accumulates from t' to t_r as:

$${\psi }_{{\rm{r}}}({t}_{{\rm{r}}})={\int }_{-\infty }^{{t}_{{\rm{r}}}}\int g({p}^{{\prime} }+{A}_{{\rm{p}}}({t}_{{\rm{r}}})){E}_{{\rm{p}}}({t}^{{\prime} })d({p}^{{\prime} }+{A}_{{\rm{p}}}({t}^{{\prime} }))\times \exp \,[-{\rm{i}}S({p}^{{\prime} },{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{p}}}(t))-{\rm{i}}\frac{{p}^{2}}{2}({t}_{{\rm{r}}}-{t}^{{\prime} })+{\rm{i}}\phi {t}^{{\prime} }]{\rm{d}}{p}^{{\prime} }{\rm{d}}{t}^{{\prime} }$$
(7)

Here d and g are the dipole and scattering matrix element, respectively, as defined in refs. [39,40,58], and $e^{i\phi t}$ reflects the extra phases acquired during the time evolution of the bound state before ionization. In our experimental setting of HAS in which the Keldysh parameter of $\gamma \approx 0.38$ suggests the tunnelling regime[60], those three processes (ionization, propagation and backscattering) can be independently treated without non-adiabatic corrections on the Volkov phase, S (refs. [61,62]).

Description of electron wave packets under addition of a weak gate field

Equation (5) implies that access to $\psi_{\rm r}(t)$ is possible if ${\widetilde{\psi }}_{{\rm{t}}}(p)$ and $A_{\rm p}(t)$ are known. Therefore our goal is to describe how these quantities can be accessed using a phase-gating process of the optical field emission by a weak replica of the driving pulse (gate pulse). Now we inspect the effects of adding the gate pulse on the dynamics of the electron described in equations (5)–(7). We define the gate pulse by a field $E_{\rm g}(t+\tau)$ and its vector potential $A_{\rm g}(t+\tau)$, in which τ is the delay between the pump and gate pulses as described above. By replacing the pump fields and its vector potentials by the superposition of pump and gate pulses in equations (5)–(7) as $E_{\rm p}(t)\rightarrow E_{\rm p}(t)+E_{\rm g}(t+\tau)$ and $A_{\rm p}(t)\rightarrow A_{\rm p}(t)+A_{\rm g}(t+\tau)$, the terminal spectral amplitude perturbed by the gate can be rewritten as:

$${\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\propto {\rm{i}}{\int }_{-\infty }^{\infty }{\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)\exp \left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\,\exp \left[-{\rm{i}}S\left(p,\infty ,{t}_{{\rm{r}}};{A}_{{\rm{p}}}(t)+{A}_{{\rm{g}}}(t+\tau)\right.\right]{\rm{d}}{t}_{{\rm{r}}}$$
(8)

in which ${\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)$ denotes the recolliding electron wave packet perturbed by the extra gate pulse as marked by the superscript (g) to be distinguished from the gate-free counterpart $\psi_{\rm r}(t)$ (see equation (7)). Because the gate-free quantity $\psi_{\rm r}(t)$ is of interest in this discussion, the subject in this section is how to express ${\widetilde{\psi }}_{{\rm{t}}}(p,\tau)$ in terms of $\psi_{\rm r}(t)$ with phase terms introduced by the gate.

First we investigate the perturbed recolliding electron wave packet ${\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)$ in equation (8) and how to link it with the gate-free electron wave packet $\psi_{\rm r}(t)$. If the gate field is sufficiently weak, that is, $\eta ={\left|{A}_{{\rm{g}}}(t)\right|}^{2}/{\left|{A}_{{\rm{p}}}(t)\right|}^{2}\ll 1$, the dipole transition and scattering amplitudes can be considered invariant, that is, $\left[{E}_{{\rm{p}}}(t)+{E}_{{\rm{g}}}(t+\tau)\right]d(p+{A}_{{\rm{p}}}(t)\,+$ ${A}_{{\rm{g}}}(t+\tau))\approx {E}_{{\rm{p}}}(t)g(p+{A}_{{\rm{p}}}(t))$ and $g(p+{A}_{{\rm{p}}}(t)+{A}_{{\rm{g}}}(t+\tau))\approx g(p+{A}_{{\rm{p}}}(t))$ in the

expression of the recolliding electron wave packet (equation (7)). In such case, the gate only modifies the phase imparted on the wave packet between ionization and recollision:

$${\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)\approx {\int }_{-\infty }^{{t}_{{\rm{r}}}}{\rm{d}}{t}^{{\prime} }\int {\rm{d}}{p}^{{\prime} }g({p}^{{\prime} }+{A}_{{\rm{p}}}({t}_{{\rm{r}}})){E}_{{\rm{p}}}({t}^{{\prime} })d({p}^{{\prime} }+{A}_{{\rm{p}}}({t}^{{\prime} }))\times \exp \,\left[-{\rm{i}}S({p}^{{\prime} },{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{p}}}(t)+{A}_{{\rm{g}}}(t+\tau))-{\rm{i}}\frac{{p}^{2}}{2}({t}_{{\rm{r}}}-{t}^{{\prime} })+{\rm{i}}\phi {t}^{{\prime} }\right]$$
(9)

In view of these considerations, the variation of the phase (see equation (6)) can be expressed as:

$$\begin{array}{l}S(p,{t}_{2},{t}_{1};{A}_{p}(t)+{A}_{g}(t+\tau))={\int }_{{t}_{1}}^{{t}_{2}}[\frac{1}{2}{[p+{A}_{p}(t)+{A}_{g}(t+\tau)]}^{2}-\frac{1}{2}{p}^{2}]{\rm{d}}t\\ \approx \mathop{\underbrace{{\int }_{{t}_{1}}^{{t}_{2}}[\frac{1}{2}{[p+{A}_{p}(t)]}^{2}-\frac{1}{2}{p}^{2}]{\rm{d}}t}}\limits_{=S(p,\,{t}_{2},{t}_{1};{A}_{p}(t))}+\mathop{\underbrace{{\int }_{{t}_{1}}^{{t}_{2}}[p+{A}_{p}(t)]{A}_{g}(t+\tau){\rm{d}}t}}\limits_{\equiv \Delta S(p+{A}_{p}(t),{t}_{2},{t}_{1};{A}_{g}(t+\tau))}\end{array}$$
(10)

Here the square term of $A_{\rm g}(t+\tau)$ is ignored, as its contribution is negligible compared with the other terms. Equation (10) implies that the gate field introduces an extra phase of $\Delta S(p+{A}_{{\rm{p}}}(t),{t}_{2},{t}_{1}\,;{A}_{{\rm{g}}}(t+\tau))$ to the gate-free case. As a result, equation (9) can be rewritten as:

$${\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)\,\approx {\int }_{-\infty }^{{t}_{{\rm{r}}}}{\rm{d}}{t}^{{\prime} }\int {\rm{d}}{p}^{{\prime} }g({p}^{{\prime} }+{A}_{{\rm{p}}}({t}_{{\rm{r}}})){E}_{{\rm{p}}}({t}^{{\prime} })d({p}^{{\prime} }+{A}_{{\rm{p}}}$$

({t}^{{\prime} }))\times \exp \left[-{\rm{i}}S({p}^{{\prime} },{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{p}}}(t))-{\rm{i}}\frac{{p}^{2}}{2}({t}_{{\rm{r}}}-{t}^{{\prime} })+{\rm{i}}\phi {t}^{{\prime} }\right]\exp \left[-{\rm{i}}\Delta S({p}^{{\prime} }+{A}_{{\rm{p}}}(t),{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{g}}}(t+\tau))\right]$$
(11)

Note that if the last phase term $e^{-i\Delta S}$ was missing, the expression would be identical to equation (7) and, thus, the gate-free recollision wave function $\psi_r(t_r)$. Hence it would obviously be convenient to remove the $e^{-i\Delta S}$ term from the integrals, as this would enable to express the recolliding electron wave packet \({\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)\) in equation (8) through the unperturbed counterpart $\psi_r(t_r)$ and an extra phase. To proceed further with this idea, we consider two approximations.

First, following the famous saddle-point approximation, the dominant contribution in the integration over the canonical momenta p' is provided by the kinetic momentum $p' + A_p(t)$ that equals the kinetic momentum p_r of the recolliding electron at the surface. Hence the kinetic momentum term in ΔS may be approximated as \(\Delta S({p}^{{\prime} }{+A}_{{\rm{p}}}(t),{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{g}}}(t+\tau))\approx \Delta S({p}_{{\rm{r}}},{t}_{{\rm{r}}},{t}^{{\prime} }{\rm{;}}{A}_{{\rm{g}}}(t+\tau))\).

Second, because the exponent term of the extra phase $e^{-i\Delta S}$ is oscillating slowly compared with e^{-iS} in the time integration over t' in equation (11), the extra phase ΔS can be approximated by a time average \(\overline{\Delta S}\) within a time window Δt:

$$\overline{\Delta S}=\frac{1}{\left|\Delta t\right|}{\int }_{{t}_{{\rm{r}}}-\Delta t}^{{t}_{{\rm{r}}}}\Delta S({p}_{{\rm{r}}},{t}_{{\rm{r}}},{t}^{{\prime} },{A}_{{\rm{g}}}(t+\tau)){\rm{d}}{t}^{{\prime} }$$
(12)

As our main goal is to reconstruct attosecond electron wave packets that contribute to the spectral cutoff, we choose Δt as the time interval between ionization and recollision of the classical backscattering trajectory that results in the highest final kinetic energy. To evaluate the averaged phase $\overline{\Delta S}$ and thereby simplify the analytical form of equation (12), ΔS can be expressed as:

$$\begin{array}{c}\Delta S({p}_{{\rm{r}}},{t}_{{\rm{r}}},{t}^{{\prime} },{A}_{{\rm{g}}}(t+\tau))={\int }_{{t}^{{\prime} }}^{{t}_{{\rm{r}}}}{p}_{{\rm{r}}}{A}_{{\rm{g}}}(t+\tau){\rm{d}}t=-{\int }_{{t}_{{\rm{r}}}}^{{\rm{\infty }}}{p}_{{\rm{r}}}{A}_{{\rm{g}}}(t+\tau){\rm{d}}t\\ \,\,\,\,\,+{\int }_{{t}^{{\prime} }}^{{\rm{\infty }}}{p}_{{\rm{r}}}{A}_{{\rm{g}}}(t+\tau){\rm{d}}t\end{array} \tag{13}$$

By inserting equation (13) into equation (12), the effective (averaged) phase variation $\overline{\Delta S}\,$ can now be evaluated:

$$\begin{array}{l}\overline{\Delta S}=-{\int }_{{t}_{r}}^{\infty }{p}_{r}{A}_{g}(t+\tau){\rm{d}}t+{\int }_{{t}_{r}}^{\infty }{p}_{r}\mathop{\underbrace{\left[\frac{1}{\Delta t}{\int }_{-\Delta t}^{0}{A}_{g}(t+t{\prime} +\tau)\,{\rm{d}}t{\prime} \right]}}\limits_{\equiv {\bar{A}}_{g}^{(b)}(t+\tau)}{\rm{d}}t\\ \,=-{\int }_{{t}_{r}}^{\infty }{p}_{r}[{A}_{g}(t+\tau)-{\bar{A}}_{g}^{(b)}(t+\tau)]\,{\rm{d}}t\\ \,=-\Delta {\rm{S}}({p}_{r},\,\infty ,{{\rm{t}}}_{{\rm{r}}};{A}_{g}(t+\tau)-{\bar{A}}_{g}^{(b)}(t+\tau))\end{array} \tag{14}$$

in which ${\bar{A}}_{{\rm{g}}}^{({\rm{b}})}(t)$ is defined as:

$${\bar{A}}_{{\rm{g}}}^{({\rm{b}})}(t)=\frac{1}{\Delta t}{\int }_{-\Delta t}^{0}{A}_{{\rm{g}}}(t+{t}^{{\prime} }){\rm{d}}{t}^{{\prime} } \tag{15}$$

Using the above-described approximations now enables to pull the extra phase term out of the integrations in equation (11) and, considering the sign flip $\overline{\Delta S}\to -\overline{\Delta S}$ at the backscattering

instance, the perturbed recolliding electron wave packet can be expressed through the gate-free wave packet and the extra averaged phase term ${\psi }_{{\rm{r}}}^{({\rm{g}})}({t}_{{\rm{r}}},\tau)\approx {\psi }_{{\rm{r}}}({t}_{{\rm{r}}}){{\rm{e}}}^{{\rm{i}}\overline{\Delta S}}$ in equation (8).

We now move on to discuss how the electron wave packet can be described at the end of the interaction in the presence of the pump and gate fields (equation (8)). Taking the results of equations (11)–(14) and restoring $p_{\rm r}$ with the kinetic momentum $p + A_{\rm p}(t)$, equation (8) can be rewritten as:

$${\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\propto {\rm{i}}{\int }_{-\infty }^{\infty }{\psi }_{{\rm{r}}}({t}_{{\rm{r}}})\exp \left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{r}\right]\exp \left[-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{p}}}(t)+{A}_{{\rm{g}}}(t+\tau))\right]\times \exp [-{\rm{i}}\Delta S(p+{A}_{{\rm{p}}}(t),\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{g}}}\left(t+\tau \right)-{\bar{A}}_{{\rm{g}}}^{({\rm{b}})}(t+\tau))]{\rm{d}}{t}_{{\rm{r}}}$$
(16)

Because the integration ranges for S and ΔS are identical (from $t_{\rm r}$ to ∞), the two phases can be merged into a single equation ($S' = S + \Delta S$),

$$\begin{array}{c}S{\prime} (p,\infty ,{t}_{r},\tau)={\int }_{{t}_{r}}^{\infty }[\frac{1}{2}{[p+{A}_{p}(t)+{A}_{g}(t+\tau)]}^{2}-\frac{1}{2}{p}^{2}]{\rm{d}}t\\ \,\,\,+\,{\int }_{{t}_{r}}^{\infty }(p+{A}_{p}(t))({A}_{g}(t+\tau)-{\bar{A}}_{g}^{(b)}(t+\tau)){\rm{d}}t\\ \,\,\,\approx \,{\int }_{{t}_{r}}^{\infty }[\frac{1}{2}{[p+{A}_{p}(t)+\mathop{\underbrace{2{A}_{g}(t+\tau)-{\bar{A}}_{g}^{(b)}(t+\tau)}}\limits_{\equiv {A}_{HAS}(t+\tau)}]}^{2}\\ \,\,\,-\,\frac{1}{2}{p}^{2}]{\rm{d}}t=S(p,\infty ,{t}_{r}\,;{A}_{p}(t)+{A}_{HAS}(t+\tau))\end{array}$$
(17)

in which $A_{\rm HAS}(t)$ is hereafter referred to as the effective HAS vector potential and reads:

$${A}_{{\rm{HAS}}}(t)=2{A}_{{\rm{g}}}(t)-{\bar{A}}_{{\rm{g}}}^{({\rm{b}})}(t)$$
(18)

This expression of the effective HAS vector potential is compatible with the classical momentum accumulation during the excursion from the ionization to the detection, $\Delta p=-\,{\rm{e}}\left[2A({t}_{{\rm{r}}})-A({t}_{{\rm{r}}}-\Delta t)\right]$, under the rescattering condition, ${\int }_{{t}_{{\rm{r}}}-\Delta t}^{{t}_{{\rm{r}}}}{A}_{{\rm{g}}}(t){\rm{d}}t=\Delta t{A}_{{\rm{g}}}({t}_{{\rm{r}}}-\Delta t)$(refs. [11,36,63]). Using equation (17), the terminal electron amplitude $({\widetilde{\psi }}_{{\rm{t}}}(p,\tau)$ can be expressed as:

$$\begin{array}{c}{\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\propto {\rm{i}}{\int }_{-\infty }^{\infty }{\psi }_{{\rm{r}}}({t}_{{\rm{r}}})\,\exp \,\left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\,\exp [-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{p}}}(t)\\ \,\,+{A}_{{\rm{HAS}}}(t+\tau))]{\rm{d}}{t}_{{\rm{r}}}\end{array}$$
(19)

Equation (19) implies that the gate also contributes to the terminal momentum of the electron wave packet by $A_{\rm HAS}(t_{\rm r} + \tau)$, which depends on the time delay τ. As described in equation (5), the momentum contribution $A_{\rm p}(t)$ of the pump field is already incorporated in the gate-free terminal electron spectral amplitude $({\widetilde{\psi }}_{{\rm{t}}}(p)$, whose intensity is directly accessible in experiments. Therefore it is convenient for the analysis of HAS data to express equation (19) with the terminal form of $({\widetilde{\psi }}_{{\rm{t}}}(p).$ We decompose the phase in equation (19) by means of $S(p,{t}_{2},{t}_{1}\,;{A}_{{\rm{p}}}(t)+{A}_{{\rm{g}}}(t+\tau))\approx S(p,{t}_{2},{t}_{1}\,;{A}_{{\rm{p}}}(t))+\Delta S(p+{A}_{{\rm{p}}}(t),{t}_{2},{t}_{1}\,;{A}_{{\rm{HAS}}}(t+\tau))$ and rewrite equation (19) as:

$${\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\propto {\rm{i}}{\int }_{-\infty }^{\infty }{\psi }_{{\rm{r}}}({t}_{{\rm{r}}})\exp \left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\exp \left[-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{p}}}(t))\right]\times \exp \left[-{\rm{i}}\Delta S(p+{A}_{{\rm{p}}}(t),\infty ,$$

{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{HAS}}}(t+\tau))\right]{\rm{d}}{t}_{{\rm{r}}}$$
(20)

In analogy to equation (11), if the $e^{-i\Delta S}$ term vanishes, the above equation is identical to equation (5), which links the recolliding electron wave packet $\psi_r(t)$ to the terminal spectral amplitude \({\widetilde{\psi }}_{{\rm{t}}}(p)\). Here, similar to the Fourier representation of the recollision wave packet, we define the Fourier pair of the terminal electron wave packet as:

$$\begin{array}{c}{\widetilde{\psi }}_{{\rm{t}}}(p)\equiv \,{\int }_{-\infty }^{\infty }{\rm{d}}t\,{\psi }_{{\rm{t}}}(t)\,\exp \left[i\frac{{p}^{2}}{2}t\right],\\ \,{\psi }_{{\rm{t}}}(t)\equiv \,{\int }_{-\infty }^{\infty }p{\rm{d}}p\,{\widetilde{\psi }}_{{\rm{t}}}(p)\,\exp \left[-i\frac{{p}^{2}}{2}t\right]\end{array}$$
(21)

Note that the terminal electron wave packet $\psi_t(t)$ is an auxiliary electron wave packet that contains time–structure information of the recolliding electron wave packet $\psi_r(t)$ at the recollision surface with the momenta translated by the Volkov propagation with the exponent \(\exp \left[-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{p}}}(t))\right]\) (see equations (1) and (5)), but without the phase from space propagation to the detection. Using the terminal electron wave packet $\psi_t(t)$ (equation (21)), the terminal electron spectral amplitude (equation (20)) can be further simplified as:

$${\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\propto {\rm{i}}{\int }_{-\infty }^{\infty }{\psi }_{{\rm{t}}}({t}_{{\rm{r}}})\exp \left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\exp \left[-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{HAS}}}(t+\tau))\right]{\rm{d}}{t}_{{\rm{r}}}$$
(22)

under the condition that the variation of the vector potential is weak during the time window of the recollision. The HAS spectrogram equation then reads:

$$I(p,\tau)={\left|{\widetilde{\psi }}_{{\rm{t}}}(p,\tau)\right|}^{2}\propto {\left|{\int }_{-\infty }^{\infty }{\psi }_{{\rm{t}}}({t}_{{\rm{r}}})\exp \left[{\rm{i}}\frac{{p}^{2}}{2}{t}_{{\rm{r}}}\right]\exp \left[-{\rm{i}}S(p,\infty ,{t}_{{\rm{r}}}{\rm{;}}{A}_{{\rm{HAS}}}(t+\tau))\right]{\rm{d}}{t}_{{\rm{r}}}\right|}^{2}$$
(23)

Equation (23) describes a spectrogram whose reconstruction allows access to the final electron wave packet $\psi_{\rm t}(t)$ and, correspondingly, ${\widetilde{\psi }}_{{\rm{t}}}(p)$ as well as $A_{\rm HAS}(t)$.

The effective HAS vector potential $A_{\rm HAS}(t)$

An explicit relationship between the incident gate vector potential $A_{\rm g}(t)$ and the effective HAS vector potential $A_{\rm HAS}(t)$ can be best understood in the Fourier domain. Using the Fourier expansion, ${A}_{{\rm{g}}}(t)={\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\rm{d}}\omega {\widetilde{A}}_{{\rm{g}}}(\omega){{\rm{e}}}^{{\rm{i}}\omega t}$, the effective HAS vector potential can be expressed as,

$${A}_{{\rm{HAS}}}(t)=2{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\widetilde{A}}_{{\rm{g}}}(\omega){{\rm{e}}}^{i\omega t}{\rm{d}}\omega -\frac{1}{\Delta t}{\int }_{-\Delta t}^{0}{\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\widetilde{A}}_{{\rm{g}}}\left(\omega \right){e}^{i\omega (t+{t}^{{\prime} })}{\rm{d}}\omega {\rm{d}}t{\prime} ={\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\widetilde{A}}_{{\rm{g}}}\left(\omega \right)\left[2-\frac{i}{\omega \Delta t}({e}^{-i\omega \Delta t}-1)\right]{e}^{i\omega t}{\rm{d}}\omega ={\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}{\widetilde{A}}_{{\rm{g}}}\left(\omega \right)\widetilde{g}(\omega){e}^{i\omega t}{\rm{d}}\omega $$
(24)

in which the newly introduced multiplier $\widetilde{g}(\omega)$ is defined as:

$$\widetilde{g}(\omega)=\left[2-\frac{{\rm{i}}}{\omega \Delta t}({{\rm{e}}}^{-{\rm{i}}\omega \Delta t}-1)\right]$$
(25)

As shown in equation (24), the Fourier components of the effective HAS vector potential ${\widetilde{A}}_{{\rm{HAS}}}(\omega)$ is related to those of the incident gate vector potential ${\widetilde{A}}_{{\rm{g}}}(\omega)$ by multiplication of $\widetilde{g}(\omega)$

$${\widetilde{A}}_{{\rm{HAS}}}(\omega)={\widetilde{A}}_{{\rm{g}}}(\omega)\widetilde{g}(\omega)$$
(26)

The gate multiplier $\widetilde{g}(\omega)$ is independent from ${\widetilde{A}}_{{\rm{g}}}(\omega)$. This allows the possibility of the complete characterization of $A_g(t)$ from $A_{HAS}(t)$ imprinted in a HAS spectrogram.

To better visualize the concept of $A_{HAS}(t)$ and to verify the validity of the assumptions used in the above derivation, a semiclassical simulation of a HAS spectrogram was performed using single-cycle pulses. The photoelectron spectrum cutoff energy variation evaluated by the HAS spectrogram is compared with the effective HAS vector potential $A_{HAS}(t)$ calculated using equation (24) (Extended Data Fig. 7). The multiplier $\widetilde{g}(\omega)$ depends on the excursion time Δt between ionization and the backscattering event of the highest-energy electron. On the basis of the well-established recollision model, 0.685 times the central excursion period[64,65] (central period of $E(\omega)/\left.{\omega }^{2}\right)$), which corresponds to about $0.85T_L$, was used for Δt to evaluate $\widetilde{g}(\omega)$. Here T_L is the centroid period of the laser pulse. Extended Data Fig. 7b–d shows that the cutoff energy variation in a HAS spectrogram closely follows $A_{HAS}(t)$ (black curve), as calculated by the unmodified vector potential of the incident pulse (dashed red curve), regardless of the carrier-envelope phase.

Retrieval of the vector potential $A_g(t)$ from a HAS spectrogram

The above discussion suggests that, by tracing the variation of the cutoff energy in a HAS spectrogram, we can obtain $A_{\mathrm{HAS}}(t)$ (red curves in Extended Data Fig. 8a,b). Therefore access to the Fourier components of the effective HAS vector potential allows the characterization of the vector potential of the incident gate: ${\widetilde{A}}_{{\rm{g}}}(\omega)={\widetilde{g}}^{-1}(\omega){\widetilde{A}}_{{\rm{HAS}}}(\omega)$ (Extended Data Fig. 8c,d). The retrieved incident vector potential $A_{\mathrm{g}}(t)$ is shown in blue in Extended Data Fig. 8b.

Identification of the absolute zero delay in a HAS spectrogram

Identification of the zero delay between pump and gate pulses in a HAS spectrogram can be obtained with various methods. Here we opted for a method that allows the absolute delay to be derived directly from the HAS spectrogram. Even though the difference between the intensities of pump and gate pulses is more than two orders of magnitude, discernible modulations (approximately 5–10%) of the spectral amplitude of the spectrogram remain. In a HAS spectrogram, the total photoelectron yield variation can be evaluated by spectral integration at each delay point (Extended Data Fig. 8e). The absolute zero-delay point can be found as the delay point at which the yield is maximally varied (vertical dashed line in Extended Data Fig. 8e).

Benchmarking HAS through EUV attosecond streaking

EUV attosecond streaking provides access to the detailed field waveform of a pulse[33,34,35]. Because this technique of field characterization is integrated in our experimental setup, it allows us to benchmark HAS as a field-characterization method.

Extended Data Fig. 9a,b shows the HAS and EUV attosecond streaking measurements, respectively. The vector potential waveform of the incident gate pulse retrieved from the cutoff analysis in HAS (red curve in Extended Data Fig. 9c) and that from EUV attosecond streaking (blue curve in Extended Data Fig. 9c) show excellent agreement, as verified by the degree

of similarity of about 0.95 (ref. [43]), and support the notion that the gate pulse indeed acts as a phase gate.

The gate pulse as a phase gate

The compact description of HAS as a spectrogram implied by equations (2) and (23) assumes that the weak replica of the pump field acts as a nearly pure phase gate on the electrons released by the pump. In other words, it can modify the momentum of electrons released by the pump field but does not greatly influence the process of electron ionization. Yet, unless the ionization nonlinearities are well understood (for instance, in atoms), a theoretical estimate of the required ratio between pump and gate pulses for attaining a sufficiently pure phase gate requires experimental validation.

To identify safe limits within which the above condition is met, we performed HAS measurements under different gate strengths and compared the vector potential waveforms extracted from HAS to those characterized by EUV attosecond streaking. As shown in Extended Data Fig. 10a,b, the vector potential waveforms from two techniques achieve best agreement at low gate/pump intensity ratio ($\eta < 10^{-2}$). At higher intensity ratios, we observe a gradually increasing disagreement between the reconstructed waveforms with the two methods (Extended Data Fig. 10c,d), implying that the gate pulse no longer serves as a weak perturbation. These measurements suggest that, for the studied system, HAS measurements require a gate pulse whose intensity is about 10^{-2} lower than the pump intensity.

HAS reconstruction methodology

At the first stage of the reconstruction of the HAS spectrogram, the terminal electron wave packet $\psi_{\rm t}(t)$ is retrieved, as its spectral intensity ${\left|{\widetilde{\psi }}_{{\rm{t}}}(p)\right|}^{2}$ can be directly obtained by a gate-free photoelectron spectrum. Therefore the reconstruction problem is reduced to retrieval of the spectral phase.

As motivated in the main text, we isolated a spectral area of interest (AOI) from 80 to 230 eV (Extended Data Fig. 11a,b). With this region of interest, the terminal wave packet can be expressed as:

$${\psi }_{{\rm{t}}}^{({\rm{AOI}})}(t)={\int }_{-{\rm{\infty }}}^{{\rm{\infty }}}\left|{\widetilde{\psi }}_{{\rm{t}}}^{({\rm{AOI}})}(\omega)\right|{{\rm{e}}}^{-{\rm{i}}{\varphi }\left({\rm{\omega }}\right)}{{\rm{e}}}^{{\rm{i}}\omega t}{\rm{d}}\omega $$
(27)

Here $\varphi(\omega)$ is the spectral phase of the electron wave packet modelled as a polynomial series up to the sixth order,

$$\varphi (\omega)=\mathop{\sum }\limits_{n}^{N=6}{D}_{n}{(\omega -{\omega }_{{\rm{c}}})}^{n}$$
(28)

in which D_n and ω_c are the nth order dispersion and central frequency, respectively. The reconstruction is based on a least-squares algorithm written in MATLAB, which aims at the total minimization of the difference among the experimental (Fig. 4a and Extended Data Fig. 11a) and reconstructed spectrogram (Fig. 4b and Extended Data Fig. 11b). To further increase the fidelity of the reconstruction, we also simultaneously fit the differential map $D(E, \tau)$ of a HAS spectrogram $I(E, \tau)$, which is defined as:

$$D(E,\tau)=\int \frac{\partial I(E,\tau)}{\partial \tau }{\rm{d}}\tau $$
(29)

The differential map is useful because it can eliminate the unmodulated intensity along the delay axis and allows the retrieval algorithm to reconstruct fine details of the experimental trace (Extended Data Fig. 11c–e). As an initial guess for the phase, zero phase was used. The retrieval of the terminal electron wave packet is shown in Extended Data Fig. 11f,g.

In a next stage of the reconstruction, the recolliding electron pulse, which is the key quantity in this work, is evaluated by the inverse Volkov propagation of the retrieved terminal electron wave packet as:

$${\psi }_{{\rm{r}}}^{({\rm{AOI}})}(t)={\int }_{-\infty }^{\infty }{\rm{d}}p\,{\widetilde{\psi }}_{{\rm{t}}}^{({\rm{AOI}})}(p)\exp \,\left[-{\rm{i}}\frac{{p}^{2}}{2}t\right]\,\exp \left[{\rm{i}}S(p,\infty ,t{\rm{;}}{A}_{{\rm{p}}}(t))\right]$$

(30)

The Volkov basis is reconstructed by precisely measuring the pump-field waveform and its timing with respect to the emission. The retrieved recolliding electron pulse is shown in Fig. 4.

Data availability

The data supporting the conclusions in the paper are available from the corresponding authors on reasonable request. Source data are provided with this paper.

References

1. Han, J.-W., Moon, D.-I. & Meyyappan, M. Nanoscale vacuum channel transistor. *Nano Lett.* **17**, 2146–2151 (2017).

2. Brodusch, N., Demers, H. & Gauvin, R. *Field Emission Scanning Electron Microscopy: New Perspectives for Materials Characterization* (Springer, 2018).

3. Hommelhoff, P. & Kling, M. *Attosecond Nanophysics: From Basic Science to Applications* (Wiley, 2015).

4. Dombi, P. et al. Strong-field nano-optics. *Rev. Mod. Phys.* **92**, 25003 (2020).

5. Hommelhoff, P., Sortais, Y., Aghajani-Talesh, A. & Kasevich, M. A. Field emission tip as a nanometer source of free electron femtosecond pulses. *Phys. Rev. Lett.* **96**, 77401 (2006).

6. Ropers, C., Solli, D. R., Schulz, C. P., Lienau, C. & Elsaesser, T. Localized multiphoton emission of femtosecond electron pulses from metal nanotips. *Phys. Rev. Lett.* **98**, 43907 (2007).

7. Krüger, M., Schenk, M. & Hommelhoff, P. Attosecond control of electrons emitted from a nanoscale metal tip. *Nature* **475**, 78–81

(2011).

8. Piglosiewicz, B. et al. Carrier-envelope phase effects on the strong-field photoemission of electrons from metallic nanostructures. *Nat. Photonics* **8**, 37–42 (2014).

9. Bormann, R., Gulde, M., Weismann, A., Yalunin, S. V. & Ropers, C. Tip-enhanced strong-field photoemission. *Phys. Rev. Lett.* **105**, 147601 (2010).

10. Niikura, H. et al. Sub-laser-cycle electron pulses for probing molecular dynamics. *Nature* **417**, 917–922 (2002).

11. Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).

12. Krausz, F. & Ivanov, M. Attosecond physics. *Rev. Mod. Phys.* **81**, 163–234 (2009).

13. Baker, S. et al. Probing proton dynamics in molecules on an attosecond time scale. *Science* **312**, 424–427 (2006).

14. Wolter, B. et al. Ultrafast electron diffraction imaging of bond breaking in di-ionized acetylene. *Science* **354**, 308–312 (2016).

15. Ferray, M. et al. Multiple-harmonic conversion of 1064 nm radiation in rare gases. *J. Phys. B At. Mol. Opt. Phys.* **21**, L31–L35 (1988).

16. Zherebtsov, S. et al. Controlled near-field enhanced electron acceleration from dielectric nanospheres with intense few-cycle laser fields. *Nat. Phys.* **7**, 656–662 (2011).

17. Herink, G., Solli, D. R., Gulde, M. & Ropers, C. Field-driven photoemission from nanostructures quenches the quiver motion. *Nature* **483**, 190–193 (2012).

18. Baum, P. & Zewail, A. H. Attosecond electron pulses for 4D diffraction and microscopy. *Proc. Natl Acad. Sci.* **104**, 18409–18414 (2007).

19. Hassan, M. T., Baskin, J. S., Liao, B. & Zewail, A. H. High-temporal-resolution electron microscopy for imaging ultrafast electron dynamics. *Nat. Photon.* **11**, 425–430 (2017).

20. Morimoto, Y. & Baum, P. Diffraction and microscopy with attosecond electron pulse trains. *Nat. Phys.* **14**, 252–256 (2018).

21. Priebe, K. E. et al. Attosecond electron pulse trains and quantum state reconstruction in ultrafast transmission electron microscopy. *Nat. Photon.* **11**, 793–797 (2017).

22. Cavalieri, A. L. et al. Attosecond spectroscopy in condensed matter. *Nature* **449**, 1029–1032 (2007).

23. Förg, B. et al. Attosecond nanoscale near-field sampling. *Nat. Commun.* **7**, 11717 (2016).

24. Dudovich, N. et al. Measuring and controlling the birth of attosecond XUV pulses. *Nat. Phys.* **2**, 781–786 (2006).

25. Shafir, D. et al. Resolving the time when an electron exits a tunnelling barrier. *Nature* **485**, 343–346 (2012).

26. Kim, K. T. et al. Manipulation of quantum paths for space–time characterization of attosecond pulses. *Nat. Phys.* **9**, 159–163 (2013).

27. Kim, K. T. et al. Petahertz optical oscilloscope. *Nat. Photon.* **7**, 958–962 (2013).

28. Van Hove, M. A., Weinberg, W. H. & Chan, C.-M. *Low-Energy Electron Diffraction: Experiment, Theory and Surface Structure Determination* Vol. 6 (Springer, 2012).

29. Schiffrin, A. et al. Optical-field-induced current in dielectrics. *Nature* **493**, 70–74 (2013).

30. Rybka, T. et al. Sub-cycle optical phase control of nanotunnelling in the single-electron regime. *Nat. Photon.* **10**, 667–670 (2016).

31. Kim, Y. H. et al. Attosecond streaking using a rescattered electron in an intense laser field. *Sci. Rep.* **10**, 22075 (2020).

32. Hoff, D. et al. Tracing the phase of focused broadband laser pulses. *Nat. Phys.* **13**, 947–951 (2017).

33. Wirth, A. et al. Synthesized light transients. *Science* **334**, 195 LP–200 (2011).

34. Hassan, M. T. et al. Optical attosecond pulses and tracking the nonlinear response of bound electrons. *Nature* **530**, 66–70 (2016).

35. Goulielmakis, E. et al. Direct measurement of light waves. *Science* **305**, 1267–1269 (2004).

36. Paulus, G. G., Becker, W., Nicklich, W. & Walther, H. Rescattering effects in above-threshold ionization: a classical model. *J. Phys. B At. Mol. Opt. Phys.* **27**, L703–L708 (1994).

37. L'Huillier, A. et al. High-order harmonic-generation cutoff. *Phys. Rev. A* **48**, R3433–R3436 (1993).

38. Lewenstein, M., Balcou, P., Ivanov, M. Y., L'Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).

39. Lewenstein, M., Kulander, K. C., Schafer, K. J. & Bucksbaum, P. H. Rings in above-threshold ionization: a quasiclassical analysis. *Phys. Rev. A* **51**, 1495–1507 (1995).

40. Milošević, D. B., Paulus, G. G., Bauer, D. & Becker, W. Above-threshold ionization by few-cycle pulses. *J. Phys. B At. Mol. Opt. Phys.* **39**, R203–R262 (2006).

41. Goulielmakis, E. et al. Single-cycle nonlinear optics. *Science* **320**, 1614–1617 (2008).

42. Hofstetter, M. et al. Attosecond dispersion control by extreme ultraviolet multilayer mirrors. *Opt. Express* **19**, 1767–1776 (2011).

43. Hecht, E. *Optics* (Pearson, 2017).

44. Tafel, A., Meier, S., Ristein, J. & Hommelhoff, P. Femtosecond laser-induced electron emission from nanodiamond-coated tungsten needle tips. *Phys. Rev. Lett.* **123**, 146802 (2019).

45. Li, C. et al. Extreme nonlinear strong-field photoemission from carbon nanotubes. *Nat. Commun.* **10**, 4891 (2019).

46. Meckel, M. et al. Laser-induced electron tunneling and diffraction. *Science* **320**, 1478–1482 (2008).

47. Garg, M. & Kern, K. Attosecond coherent manipulation of electrons in tunneling microscopy. *Science* **367**, 411–415 (2020).

48. Schultze, M. et al. State-of-the-art attosecond metrology. *J. Electron. Spectros. Relat. Phenom.* **184**, 68–77 (2011).

49. Fowler, R. H. & Nordheim, L. Electron emission in intense electric fields. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **119**, 173–181 (1928).

50. Yalunin, S. V., Gulde, M. & Ropers, C. Strong-field photoemission from surfaces: theoretical approaches. *Phys. Rev. B* **84**, 195426 (2011).

51. Rakić, A. D., Djurišić, A. B., Elazar, J. M. & Majewski, M. L. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Appl. Opt.* **37**, 5271–5283 (1998).

52. Seiffert, L., Zherebtsov, S., Kling, M. F. & Fennel, T. Strong-field physics with nanospheres. *Adv. Phys. X* **7**, 2010595 (2022).

53. Süßmann, F. et al. Field propagation-induced directionality of carrier-envelope phase-controlled photoemission from nanospheres. *Nat. Commun.* **6**, 7944 (2015).

54. Seiffert, L. et al. Attosecond chronoscopy of electron scattering in dielectric nanoparticles. *Nat. Phys.* **13**, 766–770 (2017).

55. Liu, Q. et al. Ionization-induced subcycle metallization of nanoparticles in few-cycle pulses. *ACS Photonics* **7**, 3207–3215 (2020).

56. Schötz, J. et al. Onset of charge interaction in strong-field photoemission from nanometric needle tips. *Nanophotonics* **10**, 3769–3775 (2021).

57. Seiffert, L. et al. Trapping field assisted backscattering in strong-field photoemission from dielectric nanospheres. *J. Mod. Opt.* **64**, 1096–1103 (2017).

58. Suárez, N., Chacón, A., Ciappina, M. F., Biegert, J. & Lewenstein, M. Above-threshold ionization and photoelectron spectra in atomic systems driven by strong laser fields. *Phys. Rev. A* **92**, 63421 (2015).

59. Wolkow, D. M. Über eine Klasse von Lösungen der Diracschen Gleichung. *Z. Phys.* **94**, 250–260 (1935).

60. Keldysh, L. V. Ionization in the field of a strong electromagnetic wave. *Sov. Phys. JETP* **20**, 1307–1314 (1965).

61. Pedatzur, O. et al. Attosecond tunnelling interferometry. *Nat. Phys.* **11**, 815–819 (2015).

62. Kneller, O. et al. A look under the tunnelling barrier via attosecond-gated interferometry. *Nat. Photon.* **16**, 304–310 (2022).

63. Scrinzi, A., Ivanov, M. Y., Kienberger, R. & Villeneuve, D. M. Attosecond physics. *J. Phys. B At. Mol. Opt. Phys.* **39**, R1–R37 (2005).

64. Paulus, G. G., Becker, W. & Walther, H. Classical rescattering effects in two-color above-threshold ionization. *Phys. Rev. A* **52**, 4043–4053 (1995).

65. Becker, W., Goreslavski, S. P., Milošević, D. B. & Paulus, G. G. The plateau in above-threshold ionization: the keystone of rescattering physics. *J. Phys. B At. Mol. Opt. Phys.* **51**, 162002 (2018).

Acknowledgements

We thank J. Apportin and B. Kruse for support with the FDTD simulations. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB 1477 ‘Light-Matter Interactions at Interfaces’, project number 441234705.

Funding

Open access funding provided by Universität Rostock.

Author information

Authors and Affiliations

1. Institut für Physik, Universität Rostock, Rostock, Germany

   H. Y. Kim, S. Mandal, L. Seiffert, T. Fennel & E. Goulielmakis

2. Max Planck Institute for Solid State Research, Stuttgart, Germany

   M. Garg

Contributions

H.Y.K., M.G. and S.M. performed experiments and analysed the experimental data. H.Y.K., M.G., L.S. and T.F. performed theoretical modelling and simulation. E.G. conceived and supervised the project. All authors contributed to the preparation of the manuscript.

Corresponding author

Correspondence to E. Goulielmakis.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Qing Dai, Michael Krüger and Peng Zhang for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Experimental setup.

EUV (**a**) and HAS (**b**) configurations. The schematics of the adjustable iris and the dual inner mirror assembly are reproduced with permission by SmarAct GmbH.

Extended Data Fig. 2 Module for the measurement of the absolute electron yield.

Electrons emitted from a tungsten nanotip exposed to single-cycle pulses were captured by a thin electrode placed about 2 mm above the tip. The induced voltage on the electrode was recorded by a lock-in amplifier.

Extended Data Fig. 3 Stability of electron emission from tungsten nanotips.

a, Consecutive measurements of photoelectron spectra recorded under identical conditions and over a time range of about 100 s (left). The grey line indicates the cutoff energy. Electron yield evaluated by integration of the individual spectra in the left panel (middle). Individual spectra superimposed for comparison (right). **b**, The same measurement as in **a** but over a time period of about 80 min.

Source data

Extended Data Fig. 4 Semiclassical simulations of the optical field emission from tungsten nanotips in the single-cycle limit.

a, Terminal electron energy as a function of the ionization time under the electric field of the incident single-cycle pulses (red curve). A peak field intensity of 41 TW cm^{-2} was used in accordance with the experiments. Purple and green lines indicate terminal energies of backscattered and direct electrons as functions of birth time, for which the ionization rate is encoded on the line shading in logarithmic scale. **b**, Electron spectra plotted as a function of the peak field intensity. Red and blue dashed lines indicate the high and low cutoff energies, respectively.

Source data

Extended Data Fig. 5 FDTD simulations of the near-field enhancement.

Evaluated spatial distribution of the field-enhancement factor at the apex of a tungsten nanotip with an apex radius of 35 nm and opening angle of 15° (**a**) and a similarly sized tungsten sphere (**b**).

Source data

Extended Data Fig. 6 Impact of charge interactions on electron emission dynamics.

a, Electron spectra obtained from three-dimensional semiclassical M^3C trajectory simulations for the experimental parameters without (dashed

curves) and with (solid curves) the inclusion of charge interactions for representative pulse peak intensities. **b**, Recollision energy distribution as a function of the release time without (top) and with (bottom) charge interactions (false colour denotes spectral intensity). Black and blue curves are the instantaneous energy and the temporal phase, respectively, whereas dashed and solid curves mark the cases without and with the charge interactions, respectively. **c**, The same analysis as in **b** but for the terminal energy of the electron ensemble. The charge interaction free instantaneous energy (dashed black curve) is virtually identical to that modified by charge interactions (solid black curve), except for a uniform shift (dashed grey curve) along the energy axis. **d**, Retrieved spectral intensity (blue) and phase (red) of the recolliding electron pulse with (dashed curves) and without (solid curves) inclusion of space–charge effects on the terminal electron pulse. **e**, The same analysis as in **d** but in the temporal domain.

Source data

Extended Data Fig. 7 Incident and effective HAS gate vector potentials.

a, Gate multiplier function \(\widetilde{g}(\omega)\) as defined in the text. **b**–**d**, Semiclassically simulated HAS spectrograms at different carrier-envelope phase (CEP) settings of the driving pulse. The red dashed and black solid curves indicate waveforms of the incident gate vector potential $A_g(t)$ and effective HAS gate vector potential $A_{HAS}(t)$ calculated using equation (24), respectively.

Source data

Extended Data Fig. 8 Retrieval of the vector potential of the incident gate field.

a, Measured HAS spectrogram (Fig. 3a). **b**, $A_{HAS}(t)$ (red) directly extracted by tracing the cutoff energy variation in the spectrogram and the retrieved incident vector potential $A_g(t)$ (blue). **c**, Amplitude (blue) and phase (red) of the inverse of the multiplier \(\widetilde{g}{\left(\omega \right)}^{-1}\). **d**,

Spectral intensity of $A_{HAS}(t)$ (red) and incident $A_g(t)$ (blue) vector potentials. **e**, Identification of timing between gate and pump pulses in a HAS spectrogram. The variation of electron yields is evaluated by the spectral integration of the HAS spectrogram (blue curve) for each delay. The absolute zero delay between gate and pump pulses is identified as the point at which the yield variation is maximized (dashed red line).

Source data

Extended Data Fig. 9 Benchmarking of HAS through EUV attosecond streaking.

HAS (**a**) and EUV attosecond streaking (**b**) spectrograms under identical driving-pulse field. **c**, Comparison of the vector potential of the gate pulse evaluated from HAS (red) and EUV attosecond streaking (blue) techniques.

Source data

Extended Data Fig. 10 Benchmarking HAS.

Top panels, HAS spectrograms recorded at different gate/pump intensity ratios, $\eta \approx 4.5 \times 10^{-3}$ (**a**), 8.3×10^{-3} (**b**), 2.4×10^{-3} (**c**) and 1.4×10^{-3} (**d**). Bottom panels, comparison of the vector potential waveform characterized by EUV attosecond streaking (red curve) and HAS (blue curve) for each gate/pump intensity ratio.

Source data

Extended Data Fig. 11 Retrieval of the terminal electron wave packet.

Measured (**a**) and reconstructed (**b**) HAS spectrograms in the energy range from 80 to 230 eV. Differential maps of measured (**c**) and reconstructed (**d**) spectrograms. **e**, The phase of the intensity shift in measured (blue) and reconstructed (red) spectrograms. Retrieved terminal electron wave packet in spectral (**f**) and temporal (**g**) domains. The purple and red curves denote the intensity and phase, respectively.

Source data

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Source Data Extended Data Fig. 11

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation,

distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Kim, H.Y., Garg, M., Mandal, S. *et al.* Attosecond field emission. *Nature* **613**, 662–666 (2023). https://doi.org/10.1038/s41586-022-05577-1

- Received: 25 June 2022
- Accepted: 18 November 2022
- Published: 25 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05577-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05577-1

- Article
- Open Access
- Published: 25 January 2023

A wearable cardiac ultrasound imager

- Hongjie Hu,
- Hao Huang,
- Mohan Li,
- Xiaoxiang Gao,
- Lu Yin,
- Ruixiang Qi,
- Ray S. Wu,
- Xiangjun Chen,
- Yuxiang Ma,
- Keren Shi,
- Chenghai Li,
- Timothy M. Maus,
- Brady Huang,
- Chengchangfeng Lu,
- Muyang Lin,
- Sai Zhou,
- Zhiyuan Lou,
- Yue Gu,
- Yimu Chen,
- Yusheng Lei,
- Xinyu Wang,
- Ruotao Wang,
- Wentong Yue,
- Xinyi Yang,
- Yizhou Bian,
- Jing Mu,

- Geonho Park,
- Shu Xiang,
- Shengqiang Cai,
- Paul W. Corey,
- Joseph Wang &
- …
- Sheng Xu

Show authors

Nature **volume 613**, pages 667–675 (2023)

- 20k Accesses
- 546 Altmetric
- Metrics details

Abstract

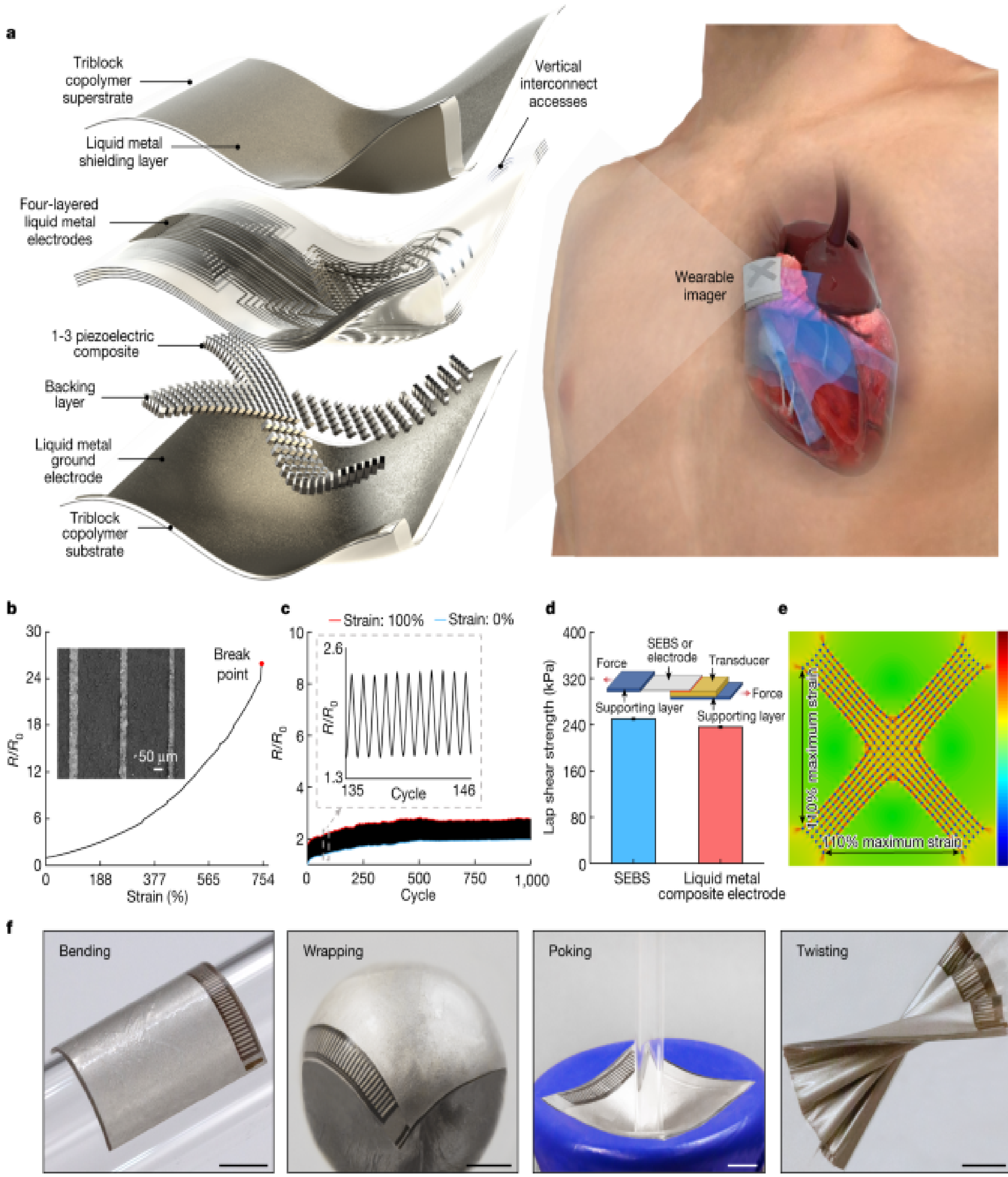
Continuous imaging of cardiac functions is highly desirable for the assessment of long-term cardiovascular health, detection of acute cardiac dysfunction and clinical management of critically ill or surgical patients[1,2,3,4]. However, conventional non-invasive approaches to image the cardiac function cannot provide continuous measurements owing to device bulkiness[5,6,7,8,9,10,11], and existing wearable cardiac devices can only capture signals on the skin[12,13,14,15,16]. Here we report a wearable ultrasonic device for continuous, real-time and direct cardiac function assessment. We introduce innovations in device design and material fabrication that improve the mechanical coupling between the device and human skin, allowing the left ventricle to be examined from different views during motion. We also develop a deep learning model that automatically extracts the left ventricular volume from the continuous image recording, yielding waveforms of key cardiac performance indices such as stroke volume, cardiac output and ejection fraction. This technology enables dynamic wearable monitoring of cardiac performance with substantially improved accuracy in various environments.

Main

Main

The device features piezoelectric transducer arrays, liquid metal composite electrodes and triblock copolymer encapsulation, as shown by the exploded schematics (Fig. 1a, left, Extended Data Fig. 1 and Supplementary Discussion 3). The device is built on styrene–ethylene–butylene–styrene (SEBS). To provide a comprehensive view of the heart, standard clinical practice is to image it in two orthogonal orientations by rotating the ultrasound probe[17]. To eliminate the need for manual rotation, we designed the device with an orthogonal configuration (Fig. 1a, right and Supplementary Videos 1 and 2). Each transducer element consisted of an anisotropic 1-3 piezoelectric composite and a silver-epoxy-based backing layer[18,19]. To balance the penetration depth and spatial resolution, we chose a centre resonant frequency of 3 MHz for deep tissue imaging[19] (Supplementary Fig. 1). The array pitch was 0.4 mm (that is, 0.78 ultrasonic wavelengths), which enhances lateral resolutions and reduces grating lobes[20].

Fig. 1: Design and characterization of the wearable cardiac imager.



a, Schematics showing the exploded view of the wearable imager, with key components labelled (left) and its working principle (right). **b**, Resistance of the liquid metal composite electrode as a function of uniaxial tensile strain. The electrode can be stretched up to about 750% without failure. The *y* axis is the relative resistance defined as R/R_0, in which R_0 and R are the measured resistances at 0% strain and a given strain, respectively. The inset is a scanning electron micrograph of the liquid metal composite electrodes with

a width as small as about 30 μm. Scale bar, 50 μm. **c**, Cycling performance of the electrode between 0% and 100% uniaxial tensile strain, showing the robustness of the electrode. The inset shows the zoomed-in features of the graph during cyclic stretching and relaxing of the electrode. **d**, Lap shear strength of the bonding between transducer elements and SEBS or liquid metal composite electrode. Data are mean and s.d. from $n = 3$ tests. The inset is a schematic setup of the lap shear test. **e**, Finite element analysis of the entire device under 110% biaxial stretching. **f**, Optical images showing the mechanical compliance of the wearable imager when bent on a developable surface, wrapped around a non-developable surface, poked and twisted. Scale bars, 5 mm.

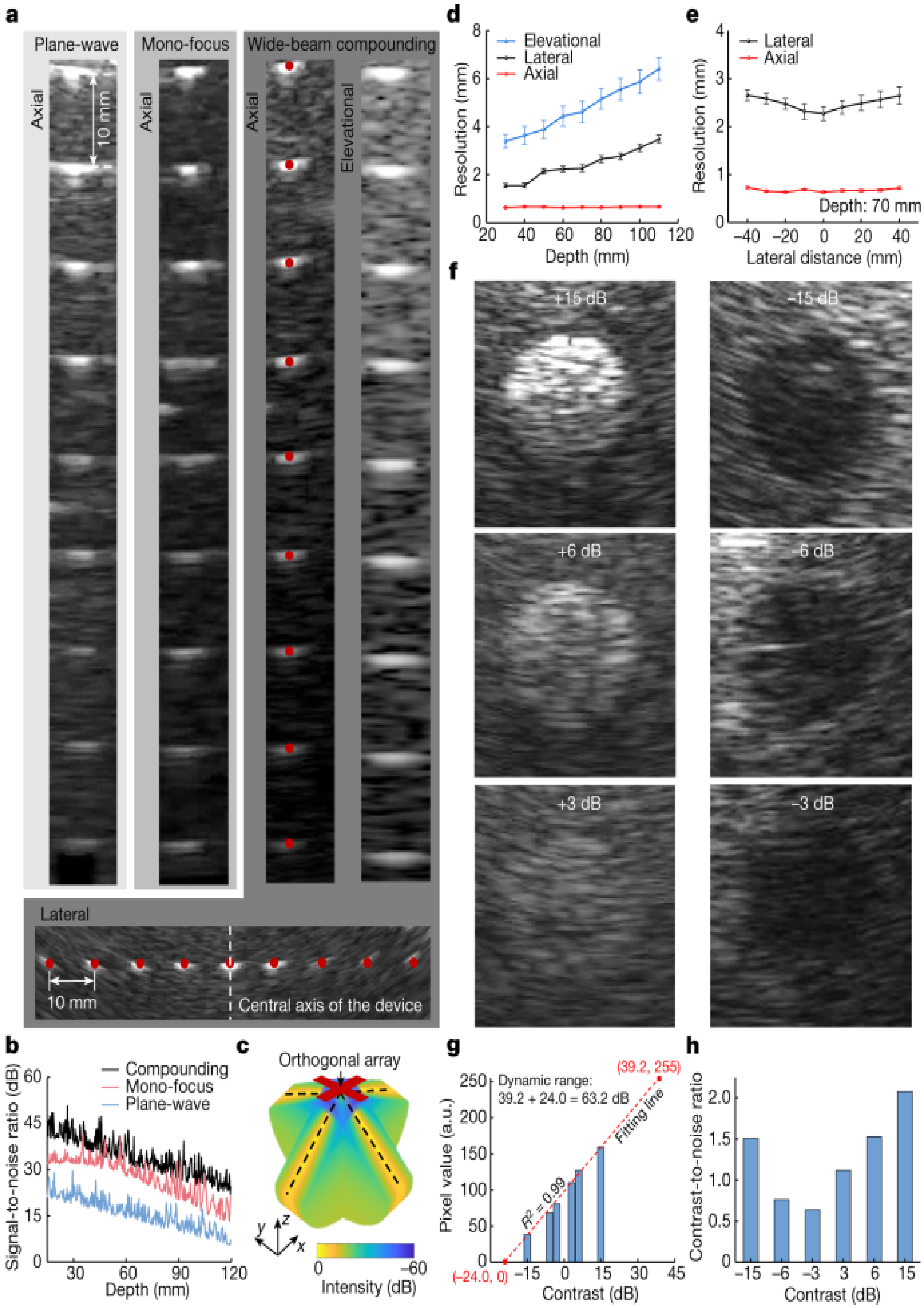
To individually address each element in such a compact array, we made high-density multilayered stretchable electrodes based on a composite of eutectic gallium–indium liquid metal and SEBS[21]. The composite is highly conductive and easy to pattern (Fig. 1b,c, Supplementary Figs. 2–4 and Methods). Lap shear measurements show that the interfacial bonding strength is about 250 kPa between the transducer element and the SEBS substrate, and about 236 kPa between the transducer element and the composite electrode (Fig. 1d and Supplementary Fig. 5), which are both stronger than typical commercial adhesives[22] (Supplementary Table 2). The resulting electrode has a thickness of only about 8 μm (Supplementary Figs. 6 and 7). Electromagnetic shielding, also made of the composite, can mitigate the interference of ambient electromagnetic waves, which reduces the noise in the ultrasound radiofrequency signals and enhances the image quality[23] (Supplementary Fig. 8 and Supplementary Discussion 4). The device has excellent electromechanical properties, as determined by its high electromechanical coupling coefficient, low dielectric loss, wide bandwidth and negligible crosstalk (Supplementary Fig. 1 and Methods). The entire device has a low Young's modulus of 921 kPa, comparable with the human skin modulus[24] (Supplementary Fig. 9). The device exhibits a high stretchability of up to approximately 110% (Fig. 1e and Supplementary Fig. 10) and can withstand various deformations (Fig. 1f). Considering that the typical strain on the human skin is within 20% (ref. [19]), these mechanical properties allow the wearable imager to maintain intimate contact with the skin over a large area, which is challenging for rigid ultrasound devices[25].

Imaging strategies and characterizations

We evaluated the quality of the generated images based on the five most crucial metrics for anatomical imaging: spatial resolutions (axial, lateral and elevational), signal-to-noise ratio, location accuracies (axial and lateral), dynamic range and contrast-to-noise-ratio[26].

The transmit beamforming strategy is critical for image quality. Therefore, we compared three distinct strategies: plane-wave, mono-focus and wide-beam compounding. Phantoms containing monofilament wires were used for this comparison (Supplementary Fig. 11, position 1). Among the three strategies, the wide-beam compounding implements a sequence of divergent acoustic waves with a series of transmission angles, and the generated images of each transmission are coherently combined to create a compounding image, which has the best quality with an expanded sonographic window[27] (Fig. 2a,b and Supplementary Figs. 12–14). We also used a receive beamforming strategy to further improve the image quality (Supplementary Fig. 15 and Methods). The wide-beam compounding achieves a synthetic focusing effect and, therefore, a high acoustic intensity across the entire insonation area (Fig. 2c and Supplementary Fig. 13), which leads to the best signal-to-noise ratio and spatial resolutions (Fig. 2a, third column, Fig. 2b and Supplementary Fig. 12).

Fig. 2: B-mode imaging strategies and characterizations.



a
Plane-wave
Mono-focus
Wide-beam compounding
Axial
10 mm
Axial
Axial
Elevational
Lateral
10 mm
Central axis of the device
b
Signal-to-noise ratio (dB)
Compounding
Mono-focus
Plane-wave
Depth (mm)
c
Orthogonal array
Intensity (dB)
d
Elevational
Lateral
Axial
Resolution (mm)
Depth (mm)
e
Lateral
Axial
Resolution (mm)
Depth: 70 mm
Lateral distance (mm)
f
+15 dB
−15 dB
+6 dB
−6 dB
+3 dB
−3 dB
g
Dynamic range:
39.2 + 24.0 = 63.2 dB
(39.2, 255)
Fitting line
R^2 = 0.99
(−24.0, 0)
Pixel value (a.u.)
Contrast (dB)
h
Contrast-to-noise ratio
Contrast (dB)

a, Imaging results on wire (100 μm in diameter) phantoms using different transmit beamforming strategies. The first three columns show the images through plane-wave, mono-focus and wide-beam compounding at different depths, respectively. The fourth column shows the imaging resolution of wide-beam compounding in the elevational direction. The bottom row shows images of laterally distributed wires by the wide-beam compounding, from which the lateral accuracy and spatial resolutions at different lateral distances from the central axis can be obtained. **b**, Signal-to-noise ratios as a function of the imaging depth under different transmission strategies. **c**, Simulated acoustic fields of the wide-beam compounding, with enhanced acoustic field across the entire insonation area. **d**, Elevational, lateral and axial resolutions of the device using wide-beam compounding at different depths. **e**, Lateral and axial resolutions of the device using wide-beam compounding with different lateral distances from the central axis. Data in **d** and **e** are mean and s.d. from five tests ($n = 5$). **f**, Imaging inclusions with different contrasts to the matrix. On the basis of these B-mode images, the dynamic range (**g**) and contrast-to-noise ratio (**h**) of the device can be quantified.

To quantify the device spatial resolutions using the wide-beam compounding strategy, we measured full widths at half maximum from the point spread function curves[28] extracted from the images (Fig. 2a, third and fourth columns and the bottom row and Supplementary Fig. 11, positions 1 and 2). As the depth increases, the elevational resolution deteriorates (Fig. 2d) because the beam becomes more divergent in the elevational direction. Therefore, we integrated six small elements into a long element (Extended Data Fig. 1) to offer better acoustic beam convergence and elevational resolution. The lateral resolution deteriorates only slightly with depth (Fig. 2d) owing to the process of receive beamforming (Methods). The axial resolution remains almost constant with depth (Fig. 2d) because it depends only on the frequency and bandwidth of the transducer array. Similarly, at the same depth, the axial resolution remains consistent with different lateral distances from the central axis of the device, whereas the lateral resolution is the best at the centre, where there is a high overlap of acoustic beams after compounding (Fig. 2e and Methods).

Another critical metric for imaging is the location accuracies. The agreements between the imaging results and the ground truths (the red dots

in Fig. 2a) in the axial and lateral directions are 96.01% and 95.90%, respectively, indicating excellent location accuracies (Methods).

Finally, we evaluated the dynamic range and contrast-to-noise ratio of the device using the wide-beam compounding strategy. Phantoms containing cylindrical inclusions with different acoustic impedances were used for the evaluation (Supplementary Fig. 11, position 3). A high acoustic impedance mismatch results in images with high contrast (Fig. 2f). We extracted the average grey values of the inclusion images and performed a linear regression[29], and determined the dynamic range to be 63.2 dB (Fig. 2g, Supplementary Fig. 16 and Methods), which is well above the 60-dB threshold typically used in medical diagnosis[30].

We selected two regions of interest, one inside and the other outside each inclusion area, to derive the contrast-to-noise ratios[31], which range from 0.63 to 2.07 (Fig. 2h and Methods). A higher inclusion contrast leads to a higher contrast-to-noise ratio of the image. The inclusions with the lowest contrast (+3 dB or −3 dB) can be clearly visualized, demonstrating the outstanding sensitivity of this device[20]. The performance of the wearable imager is comparable with that of the commercial device (Supplementary Figs. 17 and 18, Extended Data Table 1 and Supplementary Discussion 5).

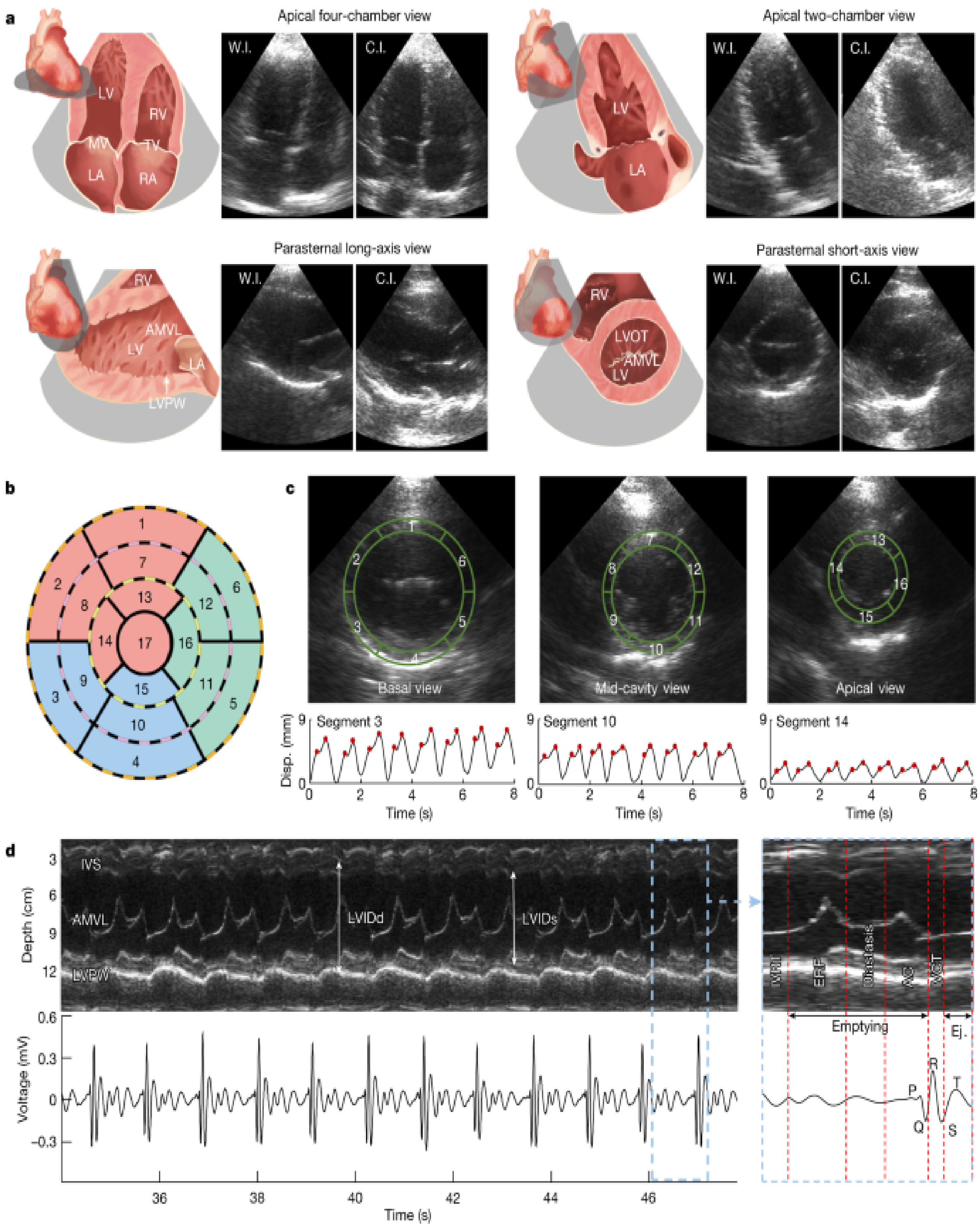
Echocardiography from several views

Echocardiography is commonly used to examine the structural integrity and blood-delivery capabilities of the heart. Uniquely for soft devices, the contours of the human chest cause a non-planar distribution of the transducer elements, which leads to phase distortion and therefore image artefacts[32]. We used a three-dimensional scanner to collect the chest curvature to compensate for element position shifts within the wearable imager and thus correct phase distortion during transmit and receive beamforming (Supplementary Fig. 19, Extended Data Fig. 2 and Supplementary Discussion 6).

We compared the performance of the wearable device with a commercial device in four primary views of echocardiography, in which critical cardiac features can be identified (Extended Data Fig. 3). Figure 3a shows the

schematics and corresponding B-mode images of these four views, including apical four-chamber view, apical two-chamber view, parasternal long-axis view and parasternal short-axis view. The difference between the results from the wearable and commercial devices is negligible. The parasternal short-axis view is particularly useful for evaluating the contractile function of the myocardium based on its motion in the radial direction and its relative thickening, as both are easily seen from this view. During contraction and relaxation, healthy myocardium undergoes strain and the wall thickness changes accordingly: thickening during contraction and thinning during relaxation. The strength of the left ventricle's contractile function can be directly reflected on the ultrasound image through the magnitude of the myocardial strain. Abnormalities in the contractile function, such as akinesia, can be indicative of ischaemic heart disease and myocardial infarction[33].

Fig. 3: Echocardiography in several standard views.



a, Schematics and B-mode images of cardiac anatomies from the wearable and commercial imagers. The wearable imager was placed in the parasternal position for imaging in the parasternal long-axis and short-axis views and relocated at the apical position for imaging in the apical four-chamber and

two-chamber views. **b**, 17-segment model representation of the left ventricular wall. Each of the concentrically nested rings that make up the circular plot represents the parasternal short-axis view of the myocardial wall from a different level of the left ventricle. **c**, B-mode images of the left ventricle in basal, mid-cavity and apical views (top row) and corresponding typical displacement for segments 3, 10 and 14, respectively (bottom row). The physical regions of the left ventricular wall represented by each segment of the 17-segment model have been labelled on the corresponding short-axis views. The peaks are marked with red dots. **d**, M-mode images (upper left) extracted from parasternal long-axis view and corresponding electrocardiogram signals (lower left). A zoomed-in plot shows the different phases of a representative cardiac cycle (right). Primary events include diastole and opening of the mitral valve during the P-wave of the electrocardiogram, opening of the aortic valve and systole during the QRS complex and closure of the aortic valve during the T-wave. AC, atrial contraction; AMVL, anterior mitral valve leaflet; C.I., commercial imager; ERF, early rapid filling; Ej., ejection; IVCT, isovolumetric contraction time; IVRT, isovolumetric relaxation time; IVS, interventricular septum; LA, left atrium; LV, left ventricle; LVIDd, left ventricular internal diameter end diastole; LVIDs, left ventricular internal diameter end systole; LVOT, left ventricular outflow tract; LVPW, left ventricular posterior wall; MV, mitral valve; RA, right atrium; RV, right ventricle; TV, tricuspid valve; W.I., wearable imager.

To better localize the specific segment of the left ventricular wall that is potentially pathological, the 17-segment model can be adopted as in standard clinical practice[33] (Fig. 3b). We took the basal, mid-cavity and apical slices of the parasternal short-axis view from the left ventricular wall, and divided them into segments according to the model. Each segment is linked to a certain coronary artery, allowing ischaemia in the coronary arteries to be localized on the basis of akinesia in the corresponding myocardial segment[33]. We then recorded the displacement waveforms of the myocardium boundaries (Fig. 3c and Supplementary Discussion 6). The two peaks in each cardiac cycle in the displacement curves correspond to the two inflows into the left ventricle during diastole. The wall displacements, as measured in the basal, mid-cavity and apical views, become sequentially

smaller owing to the decreasing radius of the myocardium along the conical shape of the left ventricle.

Motion-mode (M-mode) images track activities over time in a one-dimensional target region[34,35]. We extracted M-mode images from parasternal long-axis view B-mode images (Fig. 3d). Primary targets include the left ventricular chamber, the septum and the mitral/aortic valves. In M-mode, structural information, such as the myocardial thickness and the left ventricular diameter, can be tracked according to the distances between the boundaries of features. Valvular functions, for example, their opening and closing velocities, can be evaluated on the basis of the distance between the leaflet and septal wall (Supplementary Discussion 1). Moreover, we can correlate the mechanical activities in the M-mode images with the electrical activities in the electrocardiogram measured simultaneously during different phases in a cardiac cycle (Fig. 3d and Supplementary Discussions 1 and 6).

Monitoring during motion

Stress echocardiography assesses cardiac responses to stress induced by exercise or pharmacologic agents, which may include new or worsened ischaemia presenting as wall-motion abnormalities, and is crucial in the diagnosis of coronary artery diseases[36]. Furthermore, subjects with heart failure may sometimes seem asymptomatic at rest, as the heart sacrifices its efficiency to maintain the same cardiac output[37,38]. Thus, by pushing the heart towards its limits during exercise, the lack of efficiency becomes apparent. However, in current procedures, ultrasound images are obtained only before and after exercise. With the cumbersome apparatus, it is impossible to acquire data during exercise, which may contain invaluable real-time insights when new abnormalities initiate[39] (Supplementary Discussion 7). Also, because images are traditionally obtained after exercise, a quick recovery can mask any transient pathologic response during stress and lead to false-negative examinations[40]. In addition, the end point for terminating the exercise is subjective, which may result in suboptimal testing.

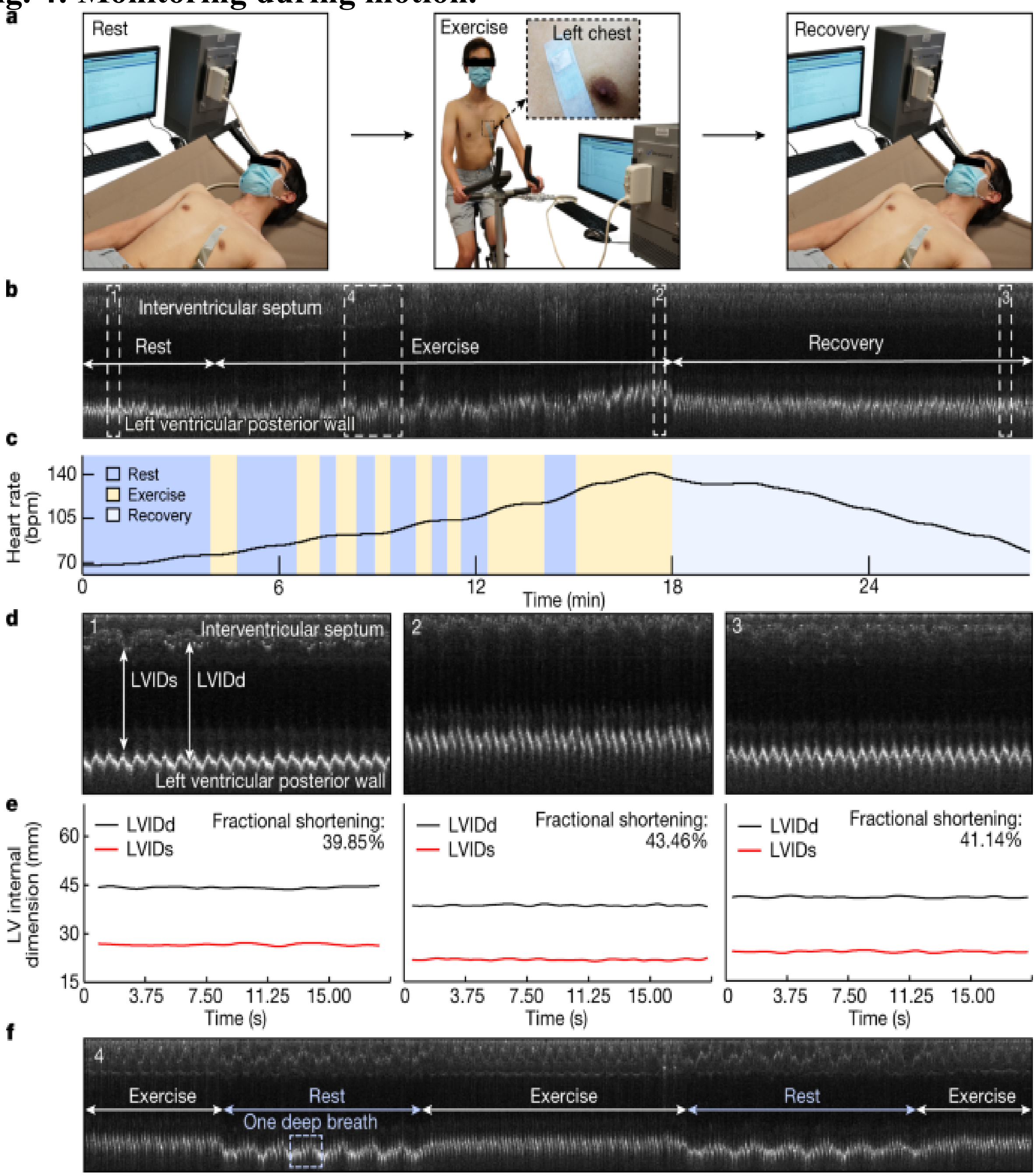
The wearable ultrasonic patch is ideal for overcoming these challenges. The device can be attached to the chest with minimal constraint to the movement

of the subject, providing a continuous recording of cardiac activities before, during and after exercise with negligible motion artefacts (Extended Data Fig. 4). This not only captures the real-time responses during the test but also offers objective data to standardize the end point and enhances patient safety throughout the test (Supplementary Discussion 7). We used liquidus silicone as the couplant to achieve images of stable quality instead of water-based ultrasound gels that evaporate over time (Supplementary Figs. 20 and 21 and Supplementary Discussion 8). We observed no skin irritation or allergy after 24 h of continuous wear (Supplementary Fig. 22). The heart rate of the subject remained stable with a constant device temperature of about 32 °C after the device continuously worked for 1 h (Supplementary Fig. 23). In addition, one device was tested on different subjects (Supplementary Fig. 24). The reproducible results indicate the stable and reliable performance of the wearable imager.

We performed stress echocardiography to demonstrate the performance of the device during exercise (Supplementary Discussion 7). The device was attached to the subject for continuous recording along the parasternal long axis during the entire process, which consisted of three main stages (Fig. 4a). In the rest stage, the subject laid in the supine position. In the exercise stage, the subject exercised on a stationary bike with several intervals until a possible maximal heart rate was reached. In the recovery stage, the subject was placed in the supine position again. The device demonstrated uninterrupted tracking of the left ventricular activities, including the corresponding M-mode echocardiography and synchronized heart-rate waveform (Fig. 4b,c, Extended Data Fig. 5 and Supplementary Video 3). We examined a representative section of each testing stage and extracted the left ventricular internal diameter end systole (LVIDs) and left ventricular internal diameter end diastole (LVIDd) (Fig. 4d). The LVIDs and LVIDd of the subject remained stable during the rest stage (Fig. 4e). In the exercise stage, the interventricular septum and left ventricular posterior wall of the subject moved closer to the skin surface, with the latter moving more than the former, resulting in a decrease in LVIDs and LVIDd. In the recovery stage, the LVIDs and LVIDd slowly returned to normal. The variation in fractional shortening, a measure of the cardiac muscular contractility, reflects the changing demand for blood supply in different stages of stress echocardiography (Fig. 4e). Particularly, section 4 in Fig. 4b includes

periods of exercise and intervals for rest, when patterns of a deep breath can also be seen from the left ventricular posterior wall motions (Fig. 4f).

Fig. 4: Monitoring during motion.



a, Three stages of stress echocardiography. In the rest stage, the subject laid supine for around 4 min. In the exercise stage, the subject rode a stationary bike for about 15 min, with intervals for rest. In the recovery stage, the subject laid supine again for about 10 min. The wearable imager was

attached to the chest of the subject throughout the entire test, even during the transitions between the stages. **b**, Continuous M-mode echocardiography extracted from the parasternal long-axis-view B-mode images of the entire process. Key features of the interventricular septum and left ventricular posterior wall are identified. The stages of rest, exercise (with intervals of rest) and recovery are labelled. **c**, Variations in the heart rate extracted from the M-mode echocardiography. **d**, Zoomed-in images of sections 1 (rest), 2 (exercise) and 3 (recovery) (dashed boxes) in **b**. **e**, Left ventricular internal diameter end diastole (LVIDd) and left ventricular internal diameter end systole (LVIDs) waveforms of the three different sections of the recording and corresponding average fractional shortenings. **f**, Zoomed-in images of section 4 (dashed box) during exercise with intervals of rest in **b**. In the first interval, the subject took a rhythmic deep breath six times, whereas during exercise, there seems to be no obvious signs of a deep breath, probably because the subject switched from diaphragmatic (rest) to thoracic (exercise) breathing, which is shallower and usually takes less time.

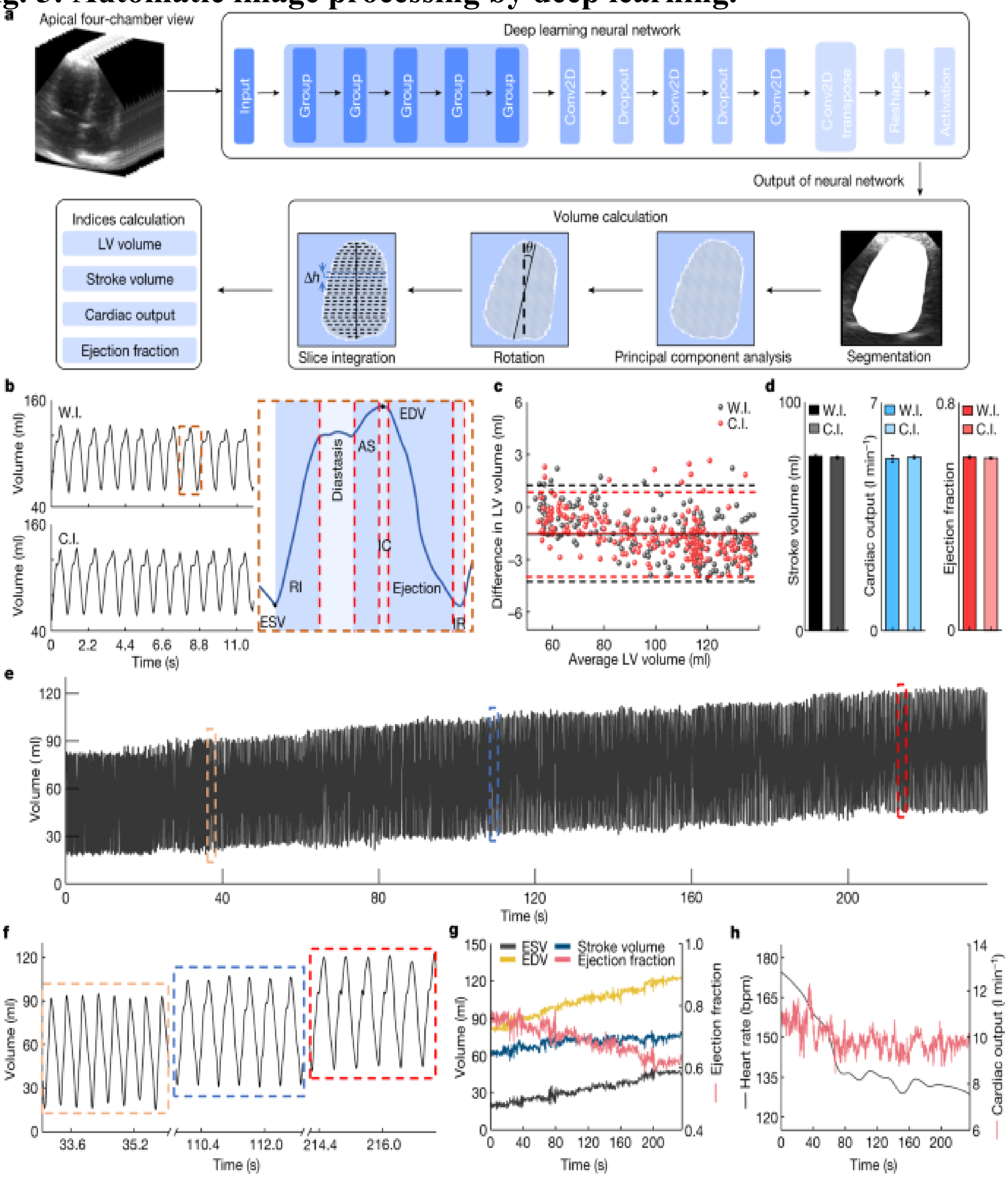
Automatic image processing

Cardiovascular diseases are often associated with changes in the pumping capabilities of the heart, which could be measured by stroke volume, cardiac output and ejection fraction. Non-invasive, continuous monitoring of these indices are valuable for the early detection and surveillance of cardiovascular conditions (Supplementary Discussion 9). Critical information embodied in these waveforms may help precisely determine potential risk factors and track the health state[41] (Supplementary Discussion 10). On the other hand, processing of the unprecedented image data streams, if done manually, can be overwhelming for clinicians, which potentially introduces interobserver variability or even errors[42].

Automatic image processing can overcome the challenges. We applied a deep learning neural network to extract key information (for example, the left ventricular volume in apical four-chamber view) from the continuous stream of images (Fig. 5a, Supplementary Fig. 25 and Supplementary Discussion 11). We evaluated different types of deep learning models[43] through the output images and waveforms of the left ventricular volume (Extended Data Figs. 6 and 7, Supplementary Table 3 and Supplementary

Video 4). The FCN-32 model outperforms others based on qualitative and quantitative analyses (Supplementary Fig. 26, Supplementary Tables 4 and 5 and Supplementary Discussion 11). We also applied data augmentation to expand the dataset and improve the performance (Supplementary Fig. 27 and Supplementary Discussion 11).

Fig. 5: Automatic image processing by deep learning.



a, Schematic workflow. Pre-processed images are used to train the FCN-32 model. The trained model accepts the unprocessed images and predicts the left ventricular (LV) volume, based on which stroke volume, cardiac output and ejection fraction are derived. **b**, Left ventricular volume waveform generated by the FCN-32 model from both the wearable imager (W.I.) and the commercial imager (C.I.) (left). Critical features are labelled in one detailed cardiac cycle (right). **c**, Bland–Altman analysis of the average of (*x* axis) and the difference between (*y* axis) the model-generated and manually labelled left ventricular volumes for the wearable (black) and commercial (red) imagers. Dashed lines indicate the 95% confidence interval and about 95% of the data points are within the interval for both imagers. Solid lines indicate mean differences. **d**, Comparing the stroke volume, cardiac output and ejection fraction extracted from results by the wearable and commercial imagers. Data are mean and s.d. from twelve cardiac cycles ($n = 12$). **e**, The model-generated left ventricular volume waveform in the recovery stage. **f**, Three representative sections of the recording from the initial, middle and end stages of **e**. End-systolic volume (ESV), end-diastolic volume (EDV), stroke volume and ejection fraction (**g**) and cardiac output and heart rate waveforms (**h**) derived from the left ventricular volume waveform. The end-systolic volume and end-diastolic volume gradually recover to the normal range in the end section. The stroke volume increases from about 60 ml to about 70 ml. The ejection fraction decreases from about 80% to about 60%. The cardiac output decreases from about 11 l min^{-1} to about 9 l min^{-1}, indicating that the decreasing heart rate from about 175 bpm to about 130 bpm overshadowed the increasing stroke volume. AS, atrial systole; IC, isovolumetric contraction; IR, isovolumetric relaxation; RI, rapid inflow.

The output left ventricular volumes for the wearable and commercial imagers show similar waveform morphologies (Fig. 5b, left). From the waveforms, corresponding phases of a cardiac cycle can be identified (Fig. 5b, right and Extended Data Fig. 8). Bland–Altman analysis gives a quantitative comparison between the model-generated and manually labelled left ventricular volumes, indicating a stable and reliable performance of the FCN-32 model[44] (Fig. 5c and Supplementary Discussion 11). The mean differences in the left ventricular volume are both approximately −1.5 ml, which is acceptable for standard medical diagnosis[45]. We then derived stroke volume, cardiac output and ejection fraction from the left ventricular volume

waveforms. No marked difference is observed in the averages or standard deviations between the two devices (Fig. 5d). The results verified the comparable performance of the wearable imager to the commercial imager.

The left ventricular volume is constantly changing and generally follows a steady-state pattern at rest in healthy subjects. Therefore, stroke volume, cardiac output and ejection fraction also tend to remain constant. However, cardiac pathologies or ordinary daily activities such as exercise may dynamically change those indices. To validate the performance of the wearable imager under dynamic situations, we extracted the left ventricular volume from recordings in the recovery stage of stress echocardiography (Fig. 5e). The dimensions of the left ventricle cannot be accurately determined when the images are collected in the standing position, owing to anatomical limitations of the human body (Supplementary Fig. 28 and Supplementary Discussion 9). Owing to the deep breathing after exercise, the heart was sometimes blocked by the lungs in the image. We used an image-imputation algorithm to complement the blocked part (Supplementary Fig. 29 and Supplementary Discussion 11). The acquired waveform shows an increasing trend in the left ventricular volume. Figure 5f illustrates three representative sections of the recording taken from the beginning, middle and end of the recovery stage. In the initial section, the diastasis stage is barely noticeable because of the high heart rate. In the middle section, the diastasis stage becomes visible. In the end section, the heart rate decreases notably. The end-diastolic and end-systolic volumes are increasing, because the slowing heartbeat during recovery allows more time for blood to fill the left ventricle[46] (Fig. 5g). The stroke volume gradually increases, indicating that the end-diastolic volume increases slightly faster than the end-systolic volume (Fig. 5g). The ejection fraction decreases as heart contraction decreases during the recovery (Fig. 5g). The cardiac output reduces, indicating a larger influence brought about by the decreasing heart rate than the increasing stroke volume (Supplementary Discussion 9).

Discussion

Echocardiography is crucial in the diagnosis of cardiac diseases, but the current implementation in clinics is cumbersome and limits its application in continuous monitoring. Emerging technologies based on wearable rigid

modules[25] or flexible patches[47] lack one or more of the ideal properties of wearable ultrasound technologies (Extended Data Table 2). In this work, we provided uninterrupted frame-by-frame acquisitions of cardiac images even when the subject was undertaking intensive exercise. In addition, the wearable imager with deep learning gave actionable information by automatically and continuously outputting curves of critical cardiac metrics, such as myocardial displacement, stroke volume, ejection fraction and cardiac output, which are highly desirable in critical care, cardiovascular disease management and sports science[48]. This capability is unprecedented in conventional clinical practice[9] and the non-invasiveness can extend potential benefits to the outpatient and athletic populations.

The implications of this technology go far beyond imaging the heart, as it can be generalized to image other deep tissues, such as the inferior vena cava, abdominal aorta, spine and liver (Supplementary Fig. 30). For example, as demonstrated in an ultrasound-guided biopsy procedure on a cyst phantom (Supplementary Fig. 31), the two orthogonal imaging sections present the entire biopsy process simultaneously, freeing up one hand of the operator (Supplementary Video 5). The uniquely enabling capability of this technology forgoes the need for an operator to constantly hold the device.

Other future efforts could ensue by further improving spatial resolutions (Supplementary Fig. 32). A three-dimensional scanner can only provide the curvature of a static human chest. To accommodate the dynamic chest curvature, advanced imaging algorithms need to be developed to compensate for the phase distortion and thus improve spatial resolutions. In addition, the wearable imager is connected to the back-end system for data processing by means of a flexible cable (Supplementary Fig. 33) and future work needs to focus on system miniaturization and integration. Besides, the FCN-32 neural network can only be applied to subjects in the training dataset at present. Its generalizability could potentially be improved by expanding the training dataset or optimizing the network with few-shot-learning[49] or reinforcement-learning[50] strategies, which will allow the model to adapt to a larger population.

Methods

Materials

Gallium–indium eutectic liquid metal, toluene, ethyl alcohol, acetone and isopropyl alcohol were purchased from Sigma-Aldrich. SEBS (G1645) was obtained from Kraton. Silicone (Ecoflex 00-30) was bought from Smooth-On as the encapsulation material of the device. Silicone (Silbione) was obtained from Elkem Silicones as the ultrasound couplant. Aquasonic ultrasound transmission gel was bought from Parker Laboratories. 1-3 composite (PZT-5H) was purchased from Del Piezo Specialties. Silver epoxy (Von Roll 3022 E-Solder) was obtained from EIS. Anisotropic conductive film cable was purchased from Elform.

Design and fabrication of the wearable imager

We designed the transducer array in an orthogonal geometry, similar to a Mills cross array (Supplementary Fig. 34), to achieve biplane standard views simultaneously. For the transducers, we chose the 1-3 composite for transmitting and receiving ultrasound waves because it possesses superior electromechanical coupling[18]. In addition, the acoustic impedance of 1-3 composites is close to that of the skin, maximizing the acoustic energy propagating into human tissues[19]. The backing layer dampens the ringing effect, broadens the bandwidth and thus improves the spatial resolution[18,51].

We used an automatic alignment strategy to fabricate the orthogonal array. The existing method of bonding the backing layer to the 1-3 composite was to first dice many small pieces of backing layer and 1-3 composite, and then bond each pair together one by one. A template was needed to align the small pieces. This method was of very low efficiency. In this study, we bond a large piece of backing layer with a large piece of 1-3 composite and then dice them together into small pieces with designed configurations. The diced array is then automatically aligned on adhesive tape with high uniformity and perfect alignment.

Electrodes based on eutectic gallium–indium liquid metal are fabricated to achieve better stretchability and higher fabrication resolution than existing electrodes based on serpentine-shaped copper thin film. Eutectic gallium–indium alloys are typically patterned through approaches such as stencil

lithography[52], masked deposition[53], inkjet printing[54], microcontact printing[55] or microfluidic channelling[56]. Although these approaches are reliable, they are either limited in patterning resolution or require sophisticated photolithography or printing hardware. The sophisticated hardware makes fabrication complicated and time-consuming, which presents a challenge in the development of compact, skin-conformal wearable electronics.

In this study, we exploited a new technology for patterning. We first screen-printed a thin layer of liquid metal on a substrate. A key consideration before screen printing was how to get the liquid metal to wet the substrate. To solve this problem, we dispersed big liquid metal particles into small microparticles using a tip sonicator (Supplementary Fig. 2). When microparticles contacted air, their outermost layer generated an oxide coating, which lowered the surface tension and prevented those microparticles from aggregating. In addition, we used 1.5 wt.% SEBS as a polymer matrix to disperse the liquid metal particles because SEBS could wet well on the liquid metal surface. We also used SEBS as the substrate. Therefore, the SEBS in the matrix and the substrate could merge and cure together after screen printing, allowing the liquid metal layer to adhere to the substrate efficiently and uniformly. Then we used laser ablation to selectively remove the liquid metal from the substrate to form patterned electrodes.

The large number of piezoelectric transducer elements in the array requires many such electrodes to address each element individually. We designed a four-layered top electrode and a common ground electrode. There are SEBS layers between different layers of liquid metal electrodes as insulation. To expose all electrode layers to connect to transducer elements, we used laser ablation to drill vertical interconnect accesses[21]. Furthermore, we created a stretchable shielding layer using liquid metal and grounded it through a vertical interconnect access, which effectively protected the device from external electromagnetic noises (Supplementary Fig. 8).

Before we attached the electrodes to the transducer array, we spin-coated toluene–ethanol solution (volume ratio 8:2) on the top of the multilayered electrode to soften the liquid-metal-based elastomer, also known as ‘solvent-welding’. The softened SEBS provided a sufficient contact surface, which

could help form a relatively strong van der Waals force between the electrodes and the metal on the transducer surface. After bonding the electrodes to the transducer array, we left the device at room temperature to let the solvent evaporate. The final bonding strength of more than 200 kPa is stronger than many commercial adhesives[22].

To encapsulate the device, we irrigated the device in a petri dish with uncured silicone elastomer (Ecoflex 00-30, Smooth-On) to fill the gap between the top and bottom electrodes and the kerf among the transducer elements. We then cured the silicone elastomer in an oven for 10 min at 80 °C. As the filling material, it suppresses spurious shear waves from adjacent elements, effectively isolating crosstalk between the elements[18,19]. With that being said, we think the main reason for the suppressed spurious shear waves is because of the epoxy in the 1-3 composite, which limits the lateral vibration of the piezoelectric materials. The Ecoflex as the filling material may have contributed but not played a chief role because the kerf is not too wide, only 100 to 200 μm. We lifted off the glass slide on the top electrode and directly covered the top electrode with a shielding layer. Then we lifted off the glass slide on the bottom electrode to release the entire device. Finally, screen-printing an approximately 50-μm layer of silicone adhesive on the device surface completed the entire fabrication.

Characterization of the liquid metal electrode

Existing wearable ultrasound arrays can achieve excellent stretchability by serpentine-shaped metal thin films as electrodes[19,26]. The serpentine geometry, however, severely limits the filling ratio of functional components, precluding the development of systems that require a high integration density or a small pitch. In this study, we chose to use liquid metal as the electrode owing to its large intrinsic stretchability, which makes the high-density electrode possible. The patterned liquid metal electrode had a minimum width of about 30 μm with a groove of about 24 μm (Supplementary Fig. 3), an order of magnitude finer than other stretchable electrodes[18,26,57]. The liquid metal electrode is ideal for connecting arrays with a small pitch[58].

This liquid metal electrode exhibited high conductivity, exceptional stretchability and negligible resistance change under tensile strain (Fig. 1b and Supplementary Fig. 4). The initial resistance at 0% strain was 1.74 Ω (corresponding to a conductivity of around 11,800 S m^{-1}), comparable with reported studies[59,60]. The resistance gradually increased with strain until the electrode reached the approximately 750% failure strain (Fig. 1b and Supplementary Fig. 4). The relative resistance is a parameter widely used to characterize the change in the resistance of a conductor (that is, the liquid metal electrode in this case) under different strains relative to the initial resistance[58,59,60]. The relative resistance is unitless. When the strain was 0%, the initial resistance R_0 was 1.74 Ω. When the electrode was under 750% strain, the electrode was broken and the resistance R at the breaking point was measured to be 44.87 Ω. Therefore, the relative resistance (R/R_0) at the breaking point was 25.79.

To investigate the electrode fatigue, we subjected them to 100% cyclic tensile strain (Fig. 1c). The initial 500 cycles observed a gradual increase in the electrode resistance because the liquid metal, when stretched, could expose more surfaces. These new surfaces were oxidized after contacting with air, leading to the resistance increase (Supplementary Fig. 4). After the initial 500 cycles, the liquid metal electrode exhibited stable resistance because, after a period of cycling, there were not many new surfaces exposed.

This study is the first to use liquid metal-based electrodes to connect ultrasound transducer elements. The bonding strength between them directly decides the robustness and endurance of the device. This is especially critical for the wearable patch, which will be subjected to repeated deformations during use. Therefore, we characterized the bonding strength of the electrode to the transducer element using a lap shear test. The liquid metal electrode was first bonded with the transducer element. The other sides of the electrode and the element were both fixed with stiff supporting layers. The supporting layer serves to be clamped by the tensile grips of the testing machine. Samples will be damaged if they are clamped by the grips directly. Then a uniaxial stretching was applied to the sample at a strain rate of 0.5 s^{-1}. The test was stopped when the electrode was delaminated from the transducer element. A SEBS film was bonded with a transducer element

and we performed the lap shear test using the same method. The peak values of the curve were used to represent the lap shear strength (Fig. 1d). The bonding strength between the pure SEBS film and the transducer element was roughly 250 kPa, and that between the electrode and the transducer element was about 236 kPa, which were both stronger than many commercial adhesives (Supplementary Table 2). The results indicate the robust bonding between the electrode and the element, preventing the electrodes from delamination under various deformations. This robust bonding does not have any limitations on the ultrasound pressures that can be transduced.

Characterization of the transducer elements

The electromechanical coupling coefficient of the transducer elements was calculated to be 0.67, on par with that of commercial probes (0.58–0.69)[61]. This superior performance was largely owing to the technique for bonding transducer elements and electrodes at room temperature in this study, which protected the piezoelectric material from heat-induced damage and depolarization. The phase angle was >60°, substantially larger than most earlier studies[18,62], indicating that most of the dipoles in the element aligned well after bonding[63]. The large phase angle also demonstrated the exceptional electromechanical coupling performance of the device. Dielectric loss is critical for evaluating the bonding process because it represents the amount of energy consumed by the transducer element at the bonding interface[20]. The average dielectric loss of the array was 0.026, on par with that of the reported rigid ultrasound probes (0.02–0.04)[64,65,66], indicating negligible energy consumed by this bonding approach (Supplementary Fig. 1b). The response echo was characterized in time and frequency domains (Supplementary Fig. 1c), from which the approximately 35 dB signal-to-noise ratio and roughly 55% bandwidth were derived. The crosstalk values between a pair of adjacent elements and a pair of second nearest neighbours have been characterized (Supplementary Fig. 1d). The average crosstalk was below the standard −30 dB in the field, indicating low mutual interference between elements.

Characterization of the wearable imager

We characterized the wearable imager using a commercial multipurpose phantom with many reflectors of different forms, layouts and acoustic impedances at various locations (CIRS ATS 539, CIRS Inc.) (Supplementary Fig. 11). The collected data are presented in Extended Data Table 1. For most of the tests, the device was first attached to the phantom surface and rotated to ensure the best imaging plane. Raw image data were saved to guarantee minimum information loss caused by the double-to-int8 conversion. Then the raw image data were processed using the ‘scanConversion’ function provided in the k-Wave toolbox to restore the sector-shaped imaging window (restored data). We applied five times upsampling in both vertical and lateral directions. The upsampled data were finally converted to the dB unit using:

$${I}_{{\rm{new}}}=20\times {\log }_{10}({I}_{{\rm{old}}})$$
(1)

The penetration depth was tested with a group of lines of higher acoustic impedance than the surrounding background distributed at different depths in the phantom. The penetration depth is defined as the depth of the deepest line that is differentiable from the background (6 dB higher in pixel value). Because the deepest line available in this study was at a depth of 16 cm and was still recognizable from the background, the penetration depth was determined as >16 cm.

The accuracy is defined as the precision of the measured distance. The accuracy was tested with the vertical and lateral groups of line phantoms. The physical distance between the two nearest pixels in the vertical and lateral directions was calculated as:

$$\Delta y=\frac{{\rm{imaging}}\,{\rm{d}}{\rm{e}}{\rm{p}}{\rm{t}}{\rm{h}}}{{N}_{{\rm{pixel}},{\rm{v}}{\rm{e}}{\rm{r}}{\rm{t}}{\rm{i}}{\rm{c}}{\rm{a}}{\rm{l}}}-1}$$
(2)
$$\Delta x=\frac{{\rm{imaging}}\,{\rm{w}}{\rm{i}}{\rm{d}}{\rm{t}}{\rm{h}}}{{N}_{{\rm{pixel}},{\rm{l}}{\rm{a}}{\rm{t}}{\rm{e}}{\rm{r}}{\rm{a}}{\rm{l}}}-1}$$
(3)

We acquired the measured distance between two lines (shown as two bright spots in the image) by counting the number of pixels between the two spots and multiplying them by Δy or Δx, depending on the measurement direction. The measured distances at different depths were compared with the ground truth described in the data sheet. Then the accuracy can be calculated by:

$${\rm{Accuracy}}=\,1-\left|\frac{{\rm{computed}}\,{\rm{d}}{\rm{i}}{\rm{s}}{\rm{t}}{\rm{a}}{\rm{n}}{\rm{c}}{\rm{e}}}{{\rm{ground}}\,{\rm{t}}{\rm{r}}{\rm{u}}{\rm{t}}{\rm{h}}}-1\right|$$
(4)

The lateral accuracy was presented as the mean accuracy of the four neighbouring pairs of lateral lines at a depth of 50 mm in the phantom.

The spatial resolutions were tested using the lateral and vertical groups of wires. For the resolutions at different depths, the full width at half maximum of the point spread function in the vertical or lateral directions for each wire was calculated. The vertical and lateral resolutions could then be derived by multiplying the number of pixels within the full width at half maximum by Δy or Δx, depending on the measurement direction. The elevational resolutions were tested by rotating the imager to form a 45° angle between the imager aperture and the lines. Then the bright spot in the B-mode images would reveal scatters out of the imaging plane. The same process as calculating the lateral resolutions was applied to obtain the elevational resolutions. The spatial resolutions at different imaging areas were also characterized with the lateral group of wires. Nine wires were located at ±4 cm, ±3 cm, ±2 cm, ±1 cm and 0 cm from the centre. The lateral and axial resolutions of the B-mode images from those wires were calculated with the same method.

Note that the lateral resolution worsens with the depth, mainly because of the receive beamforming (Supplementary Fig. 15). There are two beamformed signals, A and B. The lateral resolution of the A point (x_1) is obviously better than that of the B point (x_2). The fact that lateral resolution becomes worse with depth is inevitable in all ultrasound imaging, as long as receive beamforming is used.

As for different transmit beamforming methods, the wide-beam compounding is the best because it can achieve a synthetic focusing effect in the entire insonation area. The better the focusing effect, the higher the lateral resolution, which is why the lateral resolution of the wide-beam compounding is better than the other two transmit methods at the same depth. Furthermore, the multiple-angle scan used in the wide-beam compounding can enhance the resolution at high-angle areas. The multiple-angle scan combines transmissions at different angles to achieve a global high signal-to-noise ratio, resulting in improved resolutions.

The elevational resolution can only be characterized when the imaging target is directly beneath the transducer. For those targets that are far away from the centre, they are difficult to be imaged, which makes their elevational resolutions challenging to calculate. When characterizing the elevational resolution, the device should rotate 45°. In this case, most of the reflected ultrasound waves from those wires cannot return to the device owing to the large incidence angles. Therefore, those wires cannot be captured in the B-mode images. One potential solution is to decrease the rotating angle of the device, which may help capture more wires distributed laterally in the B-mode image. However, a small rotating angle will cause the elevational image to merge with the lateral image, which increases the error of calculating the elevational resolution. Considering those reasons, we only characterized the elevational resolution of the imaging targets directly beneath the transducer array.

The contrast resolution, the minimum contrast that can be differentiated by the imaging system, was tested with greyscale objects. The collected B-mode images are shown in Fig. 2. Because the targets with +3 and −3 dB, the lowest contrast available in this study, could still be recognized in the images, the contrast resolution of the wearable imager is determined as <3 dB.

The dynamic range in an ultrasound system refers to the contrast range that can be displayed on the monitor. The contrast between an object and the background is indicated by the average grey value of all pixels in the object in the display. The grey value is linearly proportional to the contrast. The larger the contrast, the larger the grey value. Because the display window was using the data type ‘uint8’ to differentiate the greyscale, the dynamic

range was defined as the contrast range with a grey value ranging from 0 to 255.

The object with −15 dB contrast has the lowest average grey value, whereas the object with +15 dB contrast has the highest (Supplementary Fig. 16). In our case, there are six objects with different contrasts to the background in the phantom. The highest grey value obtained from the object of +15 dB contrast was 159.8, whereas the lowest grey value from the object of −15 dB contrast was 38.7. We used a linear fit to extrapolate the contrasts when the corresponding average grey values were equal to 255 and 0, which corresponded to contrasts of 39.2 dB and −24.0 dB, respectively. Then the dynamic range was determined as:

$${\rm{Dynamic}}\,{\rm{r}}{\rm{a}}{\rm{n}}{\rm{g}}{\rm{e}}=39.2-\left(-24.0\right)=63.2\,{\rm{dB}}$$
(5)

The dead zone is defined as the depth of the first line phantom that is not overwhelmed by the initial pulses. The dead zone was tested by imaging a specific set of wire phantoms with different depths right beneath the device (Supplementary Fig. 11, position 4) directly and measuring the line phantoms that were visible in the B-mode image.

The bandwidth of the imager is defined as the ratio between the full width at half maximum in the frequency spectrum and the centre frequency. It was measured by a pulse-echo test. A piece of glass was placed 4 cm away from the device and the reflection waveform was collected with a single transducer. The collected reflection waveform was converted to the frequency spectrum by a fast Fourier transform. The full width at half maximum was read from the frequency spectrum. We obtained the bandwidth using:

$${\rm{Bandwidth}}=\frac{{\rm{full}}\,{\rm{width}}\,{\rm{at}}\,{\rm{half}}\,{\rm{maximum}}}{{\rm{centre}}\,{\rm{frequency}}}$$
(6)

Contrast sensitivity represents the capability of the device to differentiate objects with different brightness contrasts[20]. The contrast sensitivity was

tested with the greyscale objects. The contrast sensitivity is defined as the contrast-to-noise ratio (CNR) of the objects having certain contrasts to the background in the B-mode image:

$${\rm{CNR}}=\frac{\left|{\mu }_{{\rm{in}}}-{\mu }_{{\rm{out}}}\right|}{\sqrt{{\sigma }_{{\rm{in}}}^{2}+{\sigma }_{{\rm{out}}}^{2}}}$$
(7)

in which $\mu_{\rm in}$ and $\sigma_{\rm in}$ are the mean and the standard deviation of pixel intensity within the object, and $\mu_{\rm out}$ and $\sigma_{\rm out}$ are the mean and the standard deviation of pixel intensity of the background.

The insertion loss is defined as the energy loss during the transmission and receiving. It was tested in water with a quartz crystal, a function generator with an output impedance of 50 Ω and an oscilloscope (Rigol DS1104). First, the transducer received an excitation in the form of a tone burst of a 3-MHz sine wave from the function generator. Then the same transducer received the echo from the quartz crystal. Given the 1.9-dB energy loss of the transmission into the quartz crystal and the 2.2×10^{-4} dB (mm MHz)$^{-1}$ attenuation of water, the insertion loss could be calculated as:

$${\rm{Insertion}}\,{\rm{l}}{\rm{o}}{\rm{s}}{\rm{s}}=\left|20\times {\log }_{10}\left(\frac{{V}_{{\rm{r}}}}{{V}_{{\rm{t}}}}\right)+1.9+2.2\times {10}^{-4}\times 2d\times {f}_{{\rm{r}}}^{2}\right|$$
(8)

Simulation of the acoustic field

The simulation computes the root mean square of the acoustic pressure at each point in the defined simulation field. The root mean square is defined in the equation below and gives an average acoustic pressure over a certain time duration, which is pre-defined in a packaged function of the software. In the equation, x_i is the simulated acoustic pressure at the ith time step.

$${x}_{{\rm{RMS}}}=\sqrt{\frac{1}{n}({x}_{1}^{2}+{x}_{2}^{2}+\cdots +{x}_{n}^{2})}$$
(9)

Figure 2c is the simulated root mean square of the transmitted acoustic pressure field by the orthogonal transducers. The simulation was done using the MATLAB UltraSound Toolbox[67]. Each one-dimensional phased array in the orthogonal transducers gives a sector-shaped acoustic pressure field. The simulation merges two such sector-shaped acoustic pressure fields. The imaging procedure was done with the same parameters as the simulations.

In the simulation, we defined the transducer parameters first: the centre frequency of the transducers as 3 MHz, the width of the transducers as 0.3 mm, the length of the transducers as 2.3 mm, the pitch of the array as 0.4 mm, the number of elements as 32 and the bandwidth of the transducers as 55%. Then we defined wide-beam compounding (Supplementary Fig. 13) as the transmission method: 97 transmission angles, from −37.5° to +37.5°, with a step size of 0.78°. Then the acoustic pressure field was the overall effect of the 97 transmissions. Finally, we defined the computation area: −8 mm to +8 mm in the lateral direction, −6 mm to +6 mm in the elevational direction and 0 mm to 140 mm in the axial direction.

Data availability

All data are available in the manuscript or Supplementary Information.

Code availability

The code that produced the findings of this study is available at https://github.com/UCSD-XuGroup/Wearable-Cardiac-Ultrasound-Imager.

References

1. Levick, J. R. *An Introduction to Cardiovascular Physiology* (Butterworth-Heinemann, 1991).

2. Yazdanyar, A. & Newman, A. B. The burden of cardiovascular disease in the elderly: morbidity, mortality, and costs. *Clin. Geriatr. Med.* **25**, 563–577 (2009).

3. Ouyang, D. et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).

4. Jozwiak, M., Monnet, X. & Teboul, J. L. Monitoring: from cardiac output monitoring to echocardiography. *Curr. Opin. Crit. Care* **21**, 395–401 (2015).

5. Frahm, J., Voit, D. & Uecker, M. Real-time magnetic resonance imaging: radial gradient-echo sequences with nonlinear inverse reconstruction. *Invest. Radiol.* **54**, 757–766 (2019).

6. Commandeur, F., Goeller, M. & Dey, D. Cardiac CT: technological advances in hardware, software, and machine learning applications. *Curr. Cardiovasc. Imaging Rep* **11**, 19 (2018).

7. Angelidis, G. et al. SPECT and PET in ischemic heart failure. *Heart Fail. Rev.* **22**, 243–261 (2017).

8. Efimov, I. R., Nikolski, V. P. & Salama, G. Optical imaging of the heart. *Circ. Res.* **95**, 21–33 (2004).

9. Gargesha, M., Jenkins, M. W., Wilson, D. L. & Rollins, A. M. High temporal resolution OCT using image-based retrospective gating. *Opt. Express* **17**, 10786–10799 (2009).

10. Wang, R. Y. et al. High-resolution image reconstruction for portable ultrasound imaging devices. *EURASIP J. Adv. Signal Process.* **2019**, 56 (2019).

11. Baribeau, Y. et al. Handheld point-of-care ultrasound probes: the new generation of POCUS. *J. Cardiothorac. Vasc. Anesth.* **34**, 3139–3145 (2020).

12. Zimetbaum, P. J. & Josephson, M. E. Use of the electrocardiogram in acute myocardial infarction. *N. Engl. J. Med.* **348**, 933–940 (2003).

13. Alihanka, J., Vaahtoranta, K. & Saarikivi, I. A new method for long-term monitoring of the ballistocardiogram, heart rate, and respiration. *Am. J. Physiol.* **240**, R384–R392 (1981).

14. García-González, M. A., Argelagós-Palau, A., Fernández-Chimeno, M. & Ramos-Castro, J. in *Computing in Cardiology 2013* 461–464 (IEEE, 2014).

15. Elgendi, M. On the analysis of fingertip photoplethysmogram signals. *Curr. Cardiol. Rev.* **8**, 14–25 (2012).

16. Isaacson, D., Mueller, J. L., Newell, J. C. & Siltanen, S. Imaging cardiac activity by the D-bar method for electrical impedance tomography. *Physiol. Meas.* **27**, S43–S50 (2006).

17. Schiller, N. B. et al. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. *J. Am. Soc. Echocardiogr.* **2**, 358–367 (1989).

18. Hu, H. et al. Stretchable ultrasonic transducer arrays for three-dimensional imaging on complex surfaces. *Sci. Adv.* **4**, eaar3979 (2018).

19. Wang, C. et al. Monitoring of the central blood pressure waveform via a conformal ultrasonic device. *Nat. Biomed. Eng.* **2**, 687–695 (2018).

20. Shung, K. K. *Diagnostic Ultrasound: Imaging and Blood Flow Measurements* (CRC, 2005).

21. Huang, Z. L. et al. Three-dimensional integrated stretchable electronics. *Nat. Electron.* **1**, 473–480 (2018).

22. Wu, S. J., Yuk, H., Wu, J., Nabzdyk, C. S. & Zhao, X. A multifunctional origami patch for minimally invasive tissue sealing. *Adv. Mater.* **33**, e2007667 (2021).

23. Wu, H., Shen, G. & Chen, Y. A radiation emission shielding method for high intensity focus ultrasound probes. *Biomed. Mater. Eng.* **26**, S959–S966 (2015).

24. Chen, Q. P. et al. Ultrasonic inspection of curved structures with a hemispherical-omnidirectional ultrasonic probe via linear scan SAFT imaging. *NDT E Int.* **129**, 102650 (2022).

25. Wang, C. et al. Bioadhesive ultrasound for long-term continuous imaging of diverse organs. *Science* **377**, 517–523 (2022).

26. Wang, C. et al. Continuous monitoring of deep-tissue haemodynamics with stretchable ultrasonic phased arrays. *Nat. Biomed. Eng.* **5**, 749–758 (2021).

27. Montaldo, G., Tanter, M., Bercoff, J., Benech, N. & Fink, M. Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56**, 489–506 (2009).

28. Ghavami, M., Ilkhechi, A. K. & Zemp, R. Flexible transparent CMUT arrays for photoacoustic tomography. *Opt. Express* **30**, 15877–15894 (2022).

29. Xiao, Y., Boily, M., Hashemi, H. S. & Rivaz, H. High-dynamic-range ultrasound: application for imaging tendon pathology. *Ultrasound Med. Biol.* **44**, 1525–1532 (2018).

30. Zander, D. et al. Ultrasound image optimization ("knobology"): B-mode. *Ultrasound Int. Open* **6**, E14–E24 (2020).

31. Kempski, K. M., Graham, M. T., Gubbi, M. R., Palmer, T. & Lediju Bell, M. A. Application of the generalized contrast-to-noise ratio to assess photoacoustic image quality. *Biomed. Opt. Express* **11**, 3684–3698 (2020).

32. Huang, X., Lediju Bell, M. A. & Ding, K. Deep learning for ultrasound beamforming in flexible array transducer. *IEEE Trans. Med. Imaging* **40**, 3178–3189 (2021).

33. Cerqueira, M. D. et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the

Council on Clinical Cardiology of the American Heart Association. *Circulation* **105**, 539–542 (2002).

34. Feigenbaum, H. Role of M-mode technique in today's echocardiography. *J. Am. Soc. Echocardiogr.* **23**, 240–257, 335-247 (2010).

35. Devereux, R. B. et al. Standardization of M-mode echocardiographic left ventricular anatomic measurements. *J. Am. Coll. Cardiol.* **4**, 1222–1230 (1984).

36. Armstrong, W. F., Pellikka, P. A., Ryan, T., Crouse, L. & Zoghbi, W. A. Stress echocardiography: recommendations for performance and interpretation of stress echocardiography. *J. Am. Soc. Echocardiogr.* **11**, 97–104 (1998).

37. Rerych, S. K., Scholz, P. M., Newman, G. E., Sabiston, D. C. Jr & Jones, R. H. Cardiac function at rest and during exercise in normals and in patients with coronary heart disease: evaluation by radionuclide angiocardiography. *Ann. Surg.* **187**, 449–464 (1978).

38. Little, W. C. & Applegate, R. J. Congestive heart failure: systolic and diastolic function. *J. Cardiothorac. Vasc. Anesth.* **7**, 2–5 (1993).

39. Hill, J. & Timmis, A. Exercise tolerance testing. *Br. Med. J.* **324**, 1084–1087 (2002).

40. Marwick, T. H. in *Echocardiography* (eds Nihoyannopoulos, P. & Kisslo, J.) 491–519 (Springer, 2018).

41. Hammermeister, K. E., Brooks, R. C. & Warbasse, J. R. The rate of change of left ventricular volume in man: I. Validation and peak systolic ejection rate in health and disease. *Circulation* **49**, 729–738 (1974).

42. Pellikka, P. A. et al. Variability in ejection fraction measured by echocardiography, gated single-photon emission computed tomography, and cardiac magnetic resonance in patients with coronary artery disease and left ventricular dysfunction. *JAMA Netw. Open* **1**, e181456 (2018).

43. Ghorbanzadeh, O. et al. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **11**, 196 (2019).

44. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307–310 (1986).

45. Matheijssen, N. A. et al. Assessment of left ventricular volume and mass by cine magnetic resonance imaging in patients with anterior myocardial infarction intra-observer and inter-observer variability on contour detection. *Int. J. Cardiovasc. Imaging* **12**, 11–19 (1996).

46. Fritzsche, R. G., Switzer, T. W., Hodgkinson, B. J. & Coyle, E. F. Stroke volume decline during prolonged exercise is influenced by the increase in heart rate. *J. Appl. Physiol.* **86**, 799–805 (1999).

47. Pashaei, V. et al. Flexible body-conformal ultrasound patches for image-guided neuromodulation. *IEEE Trans. Biomed. Circuits Syst.* **14**, 305–318 (2020).

48. Kenny, J. S. et al. A novel, hands-free ultrasound patch for continuous monitoring of quantitative Doppler in the carotid artery. *Sci. Rep.* **11**, 7780 (2021).

49. Sung, F. et al. in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1199–1208 (IEEE, 2018).

50. Kaelbling, L. P., Littman, M. L. & Moore, A. W. Reinforcement learning: a survey. *J. Artif. Intell. Res.* **4**, 237–285 (1996).

51. Lin, M. Y., Hu, H. J., Zhou, S. & Xu, S. Soft wearable devices for deep-tissue sensing. *Nat. Rev. Mater.* **7**, 850–869 (2022).

52. Jeong, S. H. et al. Liquid alloy printing of microfluidic stretchable electronics. *Lab Chip* **12**, 4657–4664 (2012).

53. Kramer, R. K., Majidi, C. & Wood, R. J. Masked deposition of gallium-indium alloys for liquid-embedded elastomer conductors. *Adv. Funct.*

Mater. **23**, 5292–5296 (2013).

54. Ladd, C., So, J. H., Muth, J. & Dickey, M. D. 3D printing of free standing liquid metal microstructures. *Adv. Mater.* **25**, 5081–5085 (2013).

55. Tabatabai, A., Fassler, A., Usiak, C. & Majidi, C. Liquid-phase gallium–indium alloy electronics with microcontact printing. *Langmuir* **29**, 6194–6200 (2013).

56. Cheng, S. & Wu, Z. Microfluidic electronics. *Lab Chip* **12**, 2782–2791 (2012).

57. Sempionatto, J. R. et al. An epidermal patch for the simultaneous monitoring of haemodynamic and metabolic biomarkers. *Nat. Biomed. Eng.* **5**, 737–748 (2021).

58. Liu, S., Shah, D. S. & Kramer-Bottiglio, R. Highly stretchable multilayer electronic circuits using biphasic gallium-indium. *Nat. Mater.* **20**, 851–858 (2021).

59. Ma, Z. et al. Permeable superelastic liquid-metal fibre mat enables biocompatible and monolithic stretchable electronics. *Nat. Mater.* **20**, 859–868 (2021).

60. Lopes, P. A., Santos, B. C., de Almeida, A. T. & Tavakoli, M. Reversible polymer-gel transition for ultra-stretchable chip-integrated circuits through self-soldering and self-coating and self-healing. *Nat. Commun.* **12**, 4666 (2021).

61. Mi, X. H., Qin, L., Liao, Q. W. & Wang, L. K. Electromechanical coupling coefficient and acoustic impedance of 1-1-3 piezoelectric composites. *Ceram. Int.* **43**, 7374–7377 (2017).

62. Wang, Z. et al. A flexible ultrasound transducer array with micro-machined bulk PZT. *Sensors* **15**, 2538–2547 (2015).

63. Hong, C.-H. et al. Lead-free piezoceramics – where to move on? *J. Materiomics* **2**, 1–24 (2016).

64. Zhu, B. P. et al. Sol–gel derived PMN–PT thick films for high frequency ultrasound linear array applications. *Ceram. Int.* **39**, 8709–8714 (2013).

65. Li, X. et al. 80-MHz intravascular ultrasound transducer using PMN-PT free-standing film. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **58**, 2281–2288 (2011).

66. Zhu, B. et al. Lift-off PMN–PT thick film for high-frequency ultrasonic biomicroscopy. *J. Am. Ceram. Soc.* **93**, 2929–2931 (2010).

67. Shahriari, S. & Garcia, D. Meshfree simulations of ultrasound vector flow imaging using smoothed particle hydrodynamics. *Phys. Med. Biol.* **63**, 205011 (2018).

Acknowledgements

We thank Z. Wu, R. Chen and W. Zhao for guidance and discussions on experiments. We thank E. Echegaray, M. Kraushaar, X. Guo and Y. Hewei for testing and consultation of echocardiography. This work was supported by the National Institutes of Health (NIH) (1R21EB025521-01, 1R21EB027303-01A1, 3R21EB027303-02S1 and 1R01EB033464-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. All bio-experiments were conducted in accordance with the ethical guidelines of the NIH and with the approval of the Institutional Review Board of the University of California, San Diego.

Author information

Author notes

1. These authors contributed equally: Hongjie Hu, Hao Huang, Mohan Li, Xiaoxiang Gao

Authors and Affiliations

1. Department of Nanoengineering, University of California San Diego, La Jolla, CA, USA

   Hongjie Hu, Hao Huang, Xiaoxiang Gao, Lu Yin, Ray S. Wu, Yuxiang Ma, Muyang Lin, Zhiyuan Lou, Yimu Chen, Yusheng Lei, Xinyu Wang, Ruotao Wang, Wentong Yue, Yizhou Bian, Geonho Park, Joseph Wang & Sheng Xu

2. Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, USA

   Mohan Li, Chengchangfeng Lu & Sheng Xu

3. Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

   Ruixiang Qi

4. Materials Science and Engineering Program, University of California San Diego, La Jolla, CA, USA

   Xiangjun Chen, Keren Shi, Sai Zhou, Yue Gu, Xinyi Yang, Jing Mu, Shengqiang Cai, Joseph Wang & Sheng Xu

5. Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

   Yuxiang Ma

6. Materials Science and Engineering Program, University of California, Riverside, CA, USA

   Keren Shi

7. Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA, USA

   Chenghai Li & Shengqiang Cai

8. Department of Anesthesiology, University of California, San Diego Health Sulpizio Cardiovascular Center, La Jolla, CA, USA

   Timothy M. Maus

9. Department of Radiology, School of Medicine, University of California San Diego, La Jolla, CA, USA

   Brady Huang & Sheng Xu

10. Department of Neurosurgery, Yale University, New Haven, CT, USA

    Yue Gu

11. Department of Chemical Engineering, Stanford University, Stanford, CA, USA

    Yusheng Lei

12. Softsonics, Inc., San Diego, CA, USA

    Shu Xiang

13. Department of Anesthesiology, Sharp Memorial Hospital, San Diego, CA, USA

    Paul W. Corey

14. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA

    Sheng Xu

Contributions

H. Hu, H. Huang, M. Li, X.G. and S. Xu designed the research. H. Hu, H. Huang, M. Li and L.Y. performed the experiments. X.G., R.Q. and M. Li designed and trained the neural network. H. Hu, H. Huang, M. Li and Y.M. performed the data processing and simulations. H. Hu, H. Huang and S.Xu

analysed the data. H. Hu, H. Huang, M. Li, R.S.W., R.Q., S. Xiang, J.W. and S. Xu wrote the paper. All authors provided constructive and valuable feedback on the manuscript.

Corresponding author

Correspondence to Sheng Xu.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks David Ouyang, Roger Zemp and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Schematics and optical images of the orthogonal imager.

a, The orthogonal imager consists of four arms, in which six small elements in one column are combined as one long element, and a central part that is shared by the four arms. The blue and red boxes label a long element integrated by six small pieces in each direction. The number of elements in

one direction is 32. The pitch between the elements is 0.4 mm. Optical images in top view (**b**) and isometric view (**c**) showing the morphology of the orthogonal array. We used an automatic alignment strategy to fabricate the orthogonal array by bonding a large piece of backing layer with a large piece of 1-3 composite and then dicing them together into small elements with designed configurations. Inset in **c** shows the details of the elements. The 1-3 composite and backing layer have been labelled.

Extended Data Fig. 2 Characterization of the effects of phase correction on imaging quality.

B-mode images of a line phantom obtained from different situations (**a**). Left, from a planar surface. Middle, from a curvilinear surface without phase correction. Right, from a curvilinear surface with phase correction. **b**, The axial and lateral resolutions at different depths under these three situations. No obvious difference in axial resolution was found because it is mainly dependent on the transducer frequency and bandwidth. The lateral resolution of the wearable imager was improved after phase correction. Images collected in (**c**), the parasternallong-axis (PLAX) view of the heart and (**d**), the apical four-chamber view when measured by a planar probe (left panel), a curved probe without phase correction (middle panel) and a curved probe with phase correction (right panel). The left ventricular boundaries are labelled by white dashed lines in the images. **e**, Comparison of measured cardiac indices showing the impact of phase correction. Each measurement is based on the mean of five consecutive cardiac cycles (n = 5). The standard deviations are indicated by the error bars.

Extended Data Fig. 3 Optical images showing positions and orientations for ultrasound heart imaging.

a, Parasternal long-axis view. **b**, Parasternal short-axis view. **c**, Apical four-chamber view. **d**, Apical two-chamber view. The orthogonal wearable cardiac imager combines parasternal long-axis and short-axis views (**e**) and apical four-chamber and apical two-chamber views without rotation (**f**). The wearable imager can capture two parasternal views from a single position or two apical views from another single position. The sternum and ribs are labelled to indicate intercostal spaces.

Extended Data Fig. 4 B-mode images collected from a subject with different postures.

The four views collected when the subject is sitting (**a**), standing (**b**), bending over (**c**), lying flat (**d**) and lying on their side (**e**). The PLAX and PSAX views can keep their quality at different postures, whereas the quality of A4C and A2C views can only be achieved when lying on the side. A2C, apical two-chamber view; A4C, apical four-chamber view; PLAX, parasternal long-axis view; PSAX, parasternal short-axis view.

Extended Data Fig. 5 Continuous cardiac imaging during rest, exercise and recovery.

Representative B-mode and M-mode images during rest (**a**), exercise (**b**) and recovery (**c**). The red line highlights the M-mode section corresponding to the current B-mode frame. More details can be seen in Supplementary Video 3.

Extended Data Fig. 6 Segmentation results of the left ventricle with different deep learning models.

By qualitatively evaluating the result, we found no 'jitteriness' in Supplementary Video 4. The segmented left ventricle contracts and relaxes as naturally as the B-mode video. The segmentation boundaries are smooth with the highest fidelity. Compared with the original B-mode image, the FCN-32 model has the best agreement among all models used in this study.

Extended Data Fig. 7 Waveforms of the left ventricular volume obtained with different deep learning models.

Those waveforms are from segmenting the same B-mode video. Qualitatively, the waveform generated by the FCN-32 model gains the best stability and the least noise, and the waveform morphology is more constant from cycle to cycle. Quantitatively, the comparison results of those models is in Supplementary Fig. 26, which shows that the FCN-32 model has the

highest mean intersection over union, showing the best performance in this study.

Extended Data Fig. 8 Different phases in a cardiac cycle obtained from B-mode imaging.

The rows are B-mode images of A4C, A2C, PLAX and PSAX views in the same phase. The columns are B-mode images of the same view during ventricular filling, atrial contraction, isovolumetric contraction, end of ejection and isovolumetric relaxation. The dashed lines highlight the main features of the current phase. Bluish lines mean shrinking in the volume of the labelled chamber. Reddish lines mean expansion in the volume of the labelled chamber. Yellowish lines mean retention in the volume of the labelled chamber. A2C, apical two-chamber view; A4C, apical four-chamber view; LA, left atrium; LV, left ventricle; LVOT, left ventricular outflow tract; RA, right atrium; RV, right ventricle; PLAX, parasternal long-axis view; PSAX: parasternal short-axis view.

Extended Data Table 1 Full comparison of the imaging metrics between the wearable imager and a commercial ultrasound imager (Model P4-2v)
Extended Data Table 2 Summary of wearable ultrasonic devices for continuous monitoring of deep tissues

Supplementary information

Supplementary Information

This file contains Supplementary Discussions 1–11, Supplementary Figures 1–34, Supplementary Tables 1–5 and Supplementary References.

Supplementary Video 1

Cardiac long and short axis views imaged by an orthogonal array.

Supplementary Video 2

Cardiac apical four–and two–chamber views imaged by an orthogonal array.

Supplementary Video 3

Continuous cardiac imaging during rest, exercise, and recovery.

Supplementary Video 4

Left ventricle segmentation results by FCN-32.

Supplementary Video 5

Imaging guided biopsy on a phantom by an orthogonal array.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Hu, H., Huang, H., Li, M. *et al.* A wearable cardiac ultrasound imager. *Nature* **613**, 667–675 (2023). https://doi.org/10.1038/s41586-022-05498-z

- Received: 11 May 2022
- Accepted: 31 October 2022
- Published: 25 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05498-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

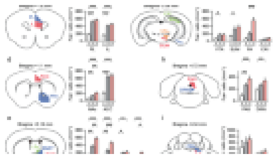
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

A wearable ultrasound patch for continuous heart imaging

Nature Research Briefing 25 Jan 2023

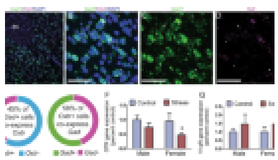
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05498-z



Social defeat stress-specific increase in c-Fos expression in the extended amygdala in mice: Involvement of dopamine D1 receptor in the medial prefrontal cortex

13 November 2019

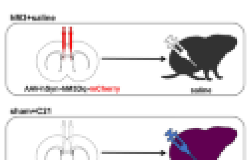
Chisato Numa, Hirotaka Nagai, … Tomoyuki Furuyashiki



Social approach and social vigilance are differentially regulated by oxytocin receptors in the nucleus accumbens

20 March 2020

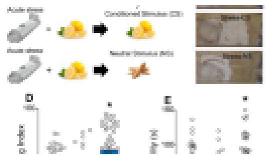
Alexia V. Williams, Natalia Duque-Wilckens, … Brian C. Trainor



Lateral septum DREADD activation alters male prairie vole prosocial and antisocial behaviors, not partner preferences

26 November 2022

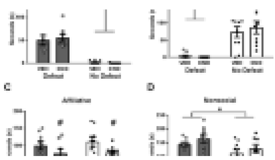
Lindsay L. Sailer, Ashley H. Park, … Alexander G. Ophir



[Behavioral and accumbens synaptic plasticity induced by cues associated with restraint stress]

05 July 2021

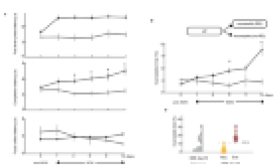
Constanza Garcia-Keller, Jordan S. Carter, … Carmela M. Reichel



[Chemogenetic activation of an infralimbic cortex to basolateral amygdala projection promotes resistance to acute social defeat stress]

23 April 2020

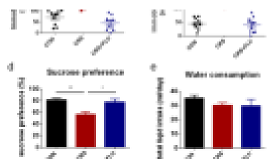
Brooke N. Dulka, Elena D. Bagatelas, … Matthew A. Cooper



[Chronic social defeat stress impairs goal-directed behavior through dysregulation of ventral hippocampal activity in male mice]

10 March 2021

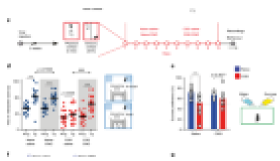
Keitaro Yoshida, Michael R. Drew, … Kenji F. Tanaka



[Stress-induced changes in social dominance are scaled by AMPA-type glutamate receptor phosphorylation in the medial prefrontal cortex]

09 October 2018

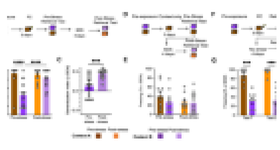
Min-Jung Park, Bo Am Seo, … Myoung-Goo Kang



[Mesopontine cholinergic inputs to midbrain dopamine neurons drive stress-induced depressive-like behaviors]

25 October 2018

Sebastian P. Fernandez, Loïc Broussot, … Jacques Barik



[Stress-induced generalization of negative memories is mediated by an extended hippocampal circuit]

07 September 2021

Lynn Y. Ren, Mariah A. A. Meyer, … Jelena Radulovic

- Article
- Open Access

- Published: 30 November 2022

Social trauma engages lateral septum circuitry to occlude social reward

- Long Li,
- Romain Durand-de Cuttoli,
- Antonio V. Aubry,
- C. Joseph Burnett,
- Flurin Cathomas,
- Lyonna F. Parise,
- Kenny L. Chan,
- Carole Morel,
- Chongzhen Yuan,
- Yusuke Shimo,
- Hsiao-yun Lin,
- Jun Wang &
- …
- Scott J. Russo

Show authors

Nature **volume 613**, pages 696–703 (2023)

- 24k Accesses
- 1 Citations
- 220 Altmetric
- Metrics details

Abstract

In humans, traumatic social experiences can contribute to psychiatric disorders[1]. It is suggested that social trauma impairs brain reward function such that social behaviour is no longer rewarding, leading to severe social avoidance[2,3]. In rodents, the chronic social defeat stress (CSDS) model has been used to understand the neurobiology underlying stress susceptibility versus resilience following social trauma, yet little is known regarding its impact on social reward[4,5]. Here we show that, following CSDS, a subset of male and female mice, termed susceptible (SUS), avoid social interaction with non-aggressive, same-sex juvenile C57BL/6J mice and do not develop context-dependent social reward following encounters with them. Non-social stressors have no effect on social reward in either sex. Next, using whole-brain Fos mapping, in vivo Ca^{2+} imaging and whole-cell recordings, we identified a population of stress/threat-responsive lateral septum neurotensin (NT^{LS}) neurons that are activated by juvenile social interactions only in SUS mice, but not in resilient or unstressed control mice. Optogenetic or chemogenetic manipulation of NT^{LS} neurons and their downstream connections modulates social interaction and social reward. Together, these data suggest that previously rewarding social targets are possibly perceived as social threats in SUS mice, resulting from hyperactive NT^{LS} neurons that occlude social reward processing.

Main

Social avoidance manifests across a host of psychiatric illnesses, with causes ranging from disinterest in social contact[6] to negative emotional states evoked by social encounters[7]. While the causes of social avoidance are diverse[8], past social trauma can result in severe social avoidance thought to reflect reduced social reward[2,9]. Despite a deep clinical understanding of social trauma and its resultant effects on social behaviour, we know very little about the underlying neural circuitry involved. Preclinical social trauma models, such as chronic social defeat stress (CSDS), have been used to better understand neural circuit mechanisms that control emotional behaviour[4,5,10]. CSDS reduces exploratory behaviours and preference for natural rewards like sucrose, and results in severe social avoidance interpreted as social anhedonia[5,10]. However, past CSDS studies assessed social interaction with an adult CD-1 mouse, similar to those used as

aggressors to induce the social trauma. Social avoidance under these circumstances probably reflects fear or submissive behaviour rather than impaired social reward.

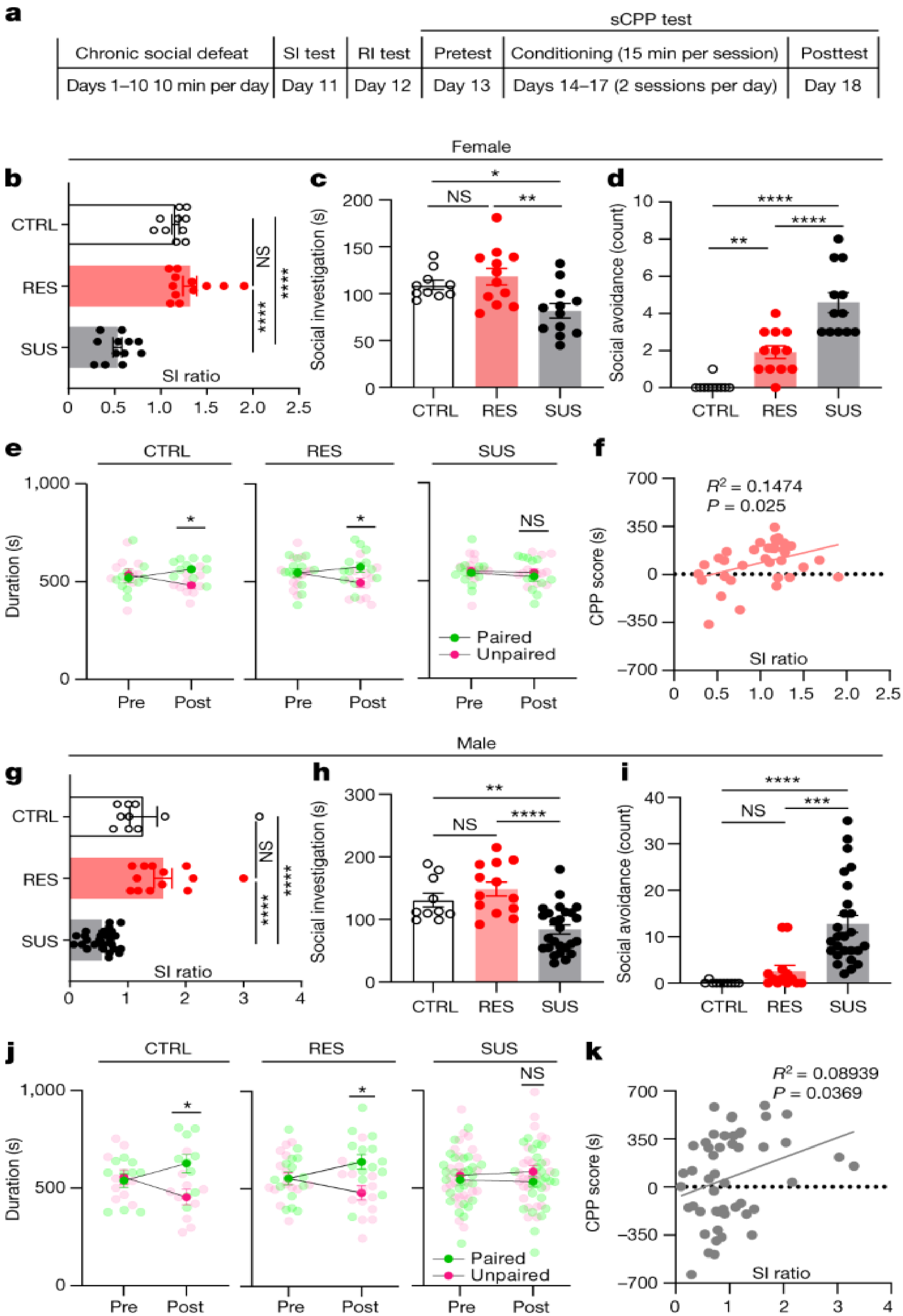
To better understand whether social reward deficits are induced by CSDS, we assessed social behaviour by testing social interaction and social conditioned place preference (sCPP) with a non-threatening, same-sex juvenile C57BL/6J mouse that, under control conditions, is rewarding. CSDS—but not non-social chronic stressors like chronic variable stress (CVS)—blocks social reward in a subset of mice termed susceptible (SUS). We next employed a circuit-based approach to better understand the mechanisms by which previous traumatic social experience with an adult male aggressor affects subsequent social reward processing. Following CSDS, SUS mice exhibit heightened activity within lateral septum neurotensin (NT^{LS}) neural circuitry, which occludes social reward and promotes sustained social avoidance behaviour even when presented with a non-threatening social situation.

SUS mice exhibit social reward deficits

To investigate how CSDS affects social interaction and social reward, 7–8-week-old wild-type (WT) male and female mice underwent standard CSDS followed by social interaction testing with a CD-1 or *ERα*-Cre F1 mouse[10,11]. As previously described, mice were classified as either resilient (RES) or SUS based on their social interaction behaviour (that is, social interaction (SI) ratio) (Fig. 1a,b,g and Extended Data Fig. 1a,b,d,e). This was followed by a resident intruder (RI) test and sCPP with 4–6-week-old, same-sex juvenile C57BL/6J mice. During the RI test, control (CTRL) and RES mice exhibited similar social behaviours towards the juvenile, including the amount of active interaction (that is, approach, close following and sniffing). Mice in these groups rarely withdrew when the juvenile approached and investigated, which we define as passive social investigation. Conversely, SUS mice exhibited much less active social investigation, longer latency to the first social bout and significantly more social avoidance during passive social investigation with a juvenile (Fig. 1c,d,h,i and Extended Data Fig. 1c,f). Social investigation time, social avoidance and latency to investigate correlated with SI ratio during testing

with a CD-1 (Extended Data Fig. 1g–l). These results show that SUS mice exhibit avoidance not only toward aggressive adult CD-1 male mice, but also toward non-threatening, same-sex C57BL/6J juvenile mice. We next used the sCPP test to assess social preference; CTRL and RES, but not SUS, mice formed social preference to the intruder-paired context (Fig. 1e,j), suggesting that juvenile interaction is not rewarding to SUS mice. sCPP score correlated with SI ratio (Fig. 1f,k) as well as social investigation time, social avoidance counts and latency to the first social bout during the RI test (Extended Data Fig. 1m–r). The female oestrous cycle was not associated with any differences in sCPP formation (Extended Data Fig. 1s). Interestingly, we found that female mice formed a significant sCPP only when the juvenile mice were confined to a wire-mesh cup during conditioning (Extended Data Fig. 1t), so we used this design for all studies in females. All behavioural parameters were normally distributed except for social avoidance (Extended Data Fig. 1u). Given that sCPP is dependent on intact learning and memory processes, we performed novel-object recognition and novel-location tests and found no evidence of learning and memory deficits in SUS or RES mice compared with CTRL mice (Extended Data Fig. 2a–c). To test whether the order in which behavioural tests were performed affected aspects of social behaviour, we reversed the order of testing (sCPP–RI–SI) in WT mice following CSDS and found similar effects (Extended Data Fig. 2d,e). Next, we grouped mice first by sCPP scores (social preference) and found a similar positive correlation with social investigation in the RI test along with a trend for SI ratio (Extended Data Fig. 2f,g), which again suggests that these different social behaviours largely correlate with one another. Together, these data support the idea that CSDS-induced social avoidance results from disruptions in social reward processing, which led us to consider that SUS mice may in fact perceive juvenile social targets as threatening or stressful.

Fig. 1: Susceptible mice show social reward impairment after CSDS in both males and females.


a
sCPP test
Chronic social defeat
SI test
RI test
Pretest
Conditioning (15 min per session)
Posttest
Days 1–10 10 min per day
Day 11
Day 12
Day 13
Days 14–17 (2 sessions per day)
Day 18
Female
b
CTRL
RES
SUS
SI ratio
NS
****
****
c
Social investigation (s)
CTRL
RES
SUS
d
Social avoidance (count)
CTRL
RES
SUS
e
CTRL
RES
SUS
Duration (s)
Pre
Post
Paired
Unpaired
f
R^2 = 0.1474
P = 0.025
CPP score (s)
SI ratio
Male
g
CTRL
RES
SUS
SI ratio
h
Social investigation (s)
i
Social avoidance (count)
j
CTRL
RES
SUS
Duration (s)
Pre
Post
Paired
Unpaired
k
R^2 = 0.08939
P = 0.0369
CPP score (s)
SI ratio

a, Experimental timeline for social behaviour tests following chronic social defeat. **b**,**g**, SI ratios of females (one-way analysis of variance (ANOVA) with Tukey's multiple-comparisons test, F (2, 31) = 53.96, $P < 0.0001$, n = 10 (CTRL), 12 (RES), 12 (SUS) (**b**); and of males (one-way ANOVA, F (2, 46) = 24.36, $P < 0.0001$, n = 10 (CTRL), 13 (RES), 26 (SUS) (**g**). **c**,**h**, Social investigation time from RI test of females (one-way ANOVA, F (2, 31) = 6.755, $P = 0.0037$) (**c**); and of males (F (2, 46) = 14.82, $P < 0.0001$) (**h**). **d**,**i**, Social avoidance of females (F (2, 31) = 33.13, $P < 0.0001$) (**d**) and males (F (2, 46) = 15.37, $P < 0.0001$) (**i**). **e**, Time spent in paired and unpaired chambers during sCPP test (female CTRL (two-way repeated-measures ANOVA followed by Šídák's multiple-comparisons test, F (1, 18) = 7.023, $P = 0.0163$, $n = 10$); RES mice (F (1, 22) = 4.598, $P = 0.0433$, $n = 12$); and SUS mice (F (1, 22) = 0.08155, $P = 0.7779$, $n = 12$)). **f**, Correlation between SI ratio and sCPP in females ($R^2 = 0.1474$, $P = 0.025$). **j**, Time spent in paired and unpaired chambers during sCPP test (two-way repeated-measures ANOVA: male CTRL (F (1, 18) = 6.074, $P = 0.0240$, $n = 10$); RES mice (F (1, 26) = 7.499, $P = 0.0110$, $n = 13$); and SUS mice (F (1, 50) = 0.4818, $P = 0.4908$, $n = 26$)). **k**, Correlation between SI ratio and sCPP in males ($R^2 = 0.08939$, $P = 0.0369$). NS, not significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. All data expressed as mean ± s.e.m.

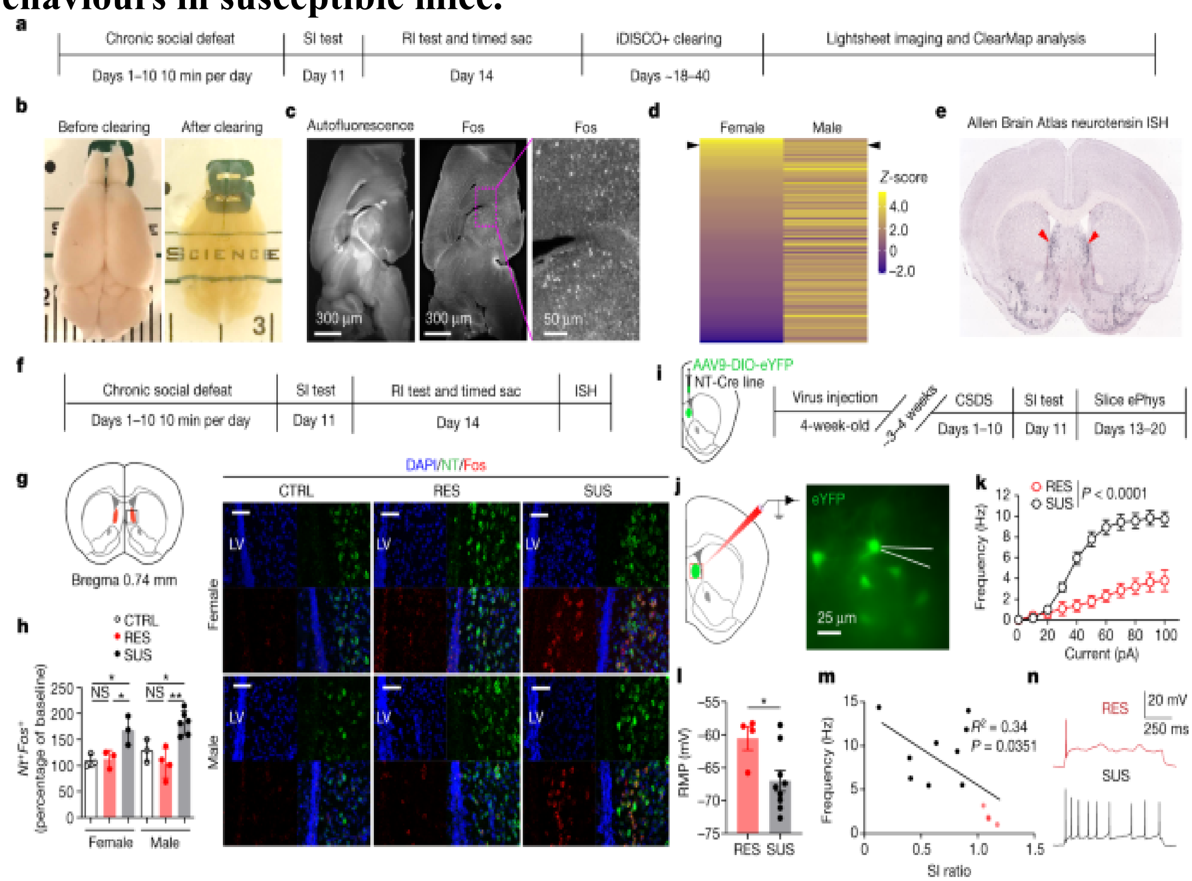
Source data

NT^{LS} neurons are hyperactivated in SUS mice

To investigate the circuit mechanisms mediating social reward deficits in SUS mice, we conducted a cleared whole-brain Fos mapping procedure using the iDISCO+ method[12] to examine differentially modulated brain regions following CSDS when mice were exposed to juvenile intruders (Fig. 2a and Extended Data Fig. 3a–h). Cleared brains (Fig. 2b) were imaged on a lightsheet microscope (Fig. 2c), followed by registration and annotation (Extended Data Fig. 3i) using ClearMap[12]. To first screen potentially relevant brain regions, Fos^+ cells were compared among CTRL, SUS and RES mice to identify differentially regulated brain regions in both sexes (Fig. 2d, Extended Data Tables 1–5 and Supplementary Table 1). Interestingly, we found dramatic sex-based differences in Fos activity when

comparing RES and SUS mice, despite both sexes exhibiting similar social deficits. Compared with RES males, SUS males showed a significant increase in Fos^+ cells in 47 regions whereas SUS females showed significant increases in only 22 regions. Notably, the lateral septum (LS) was one of the most highly activated brain regions in both SUS males and females compared with RES mice, so we selected it for further investigation. To confirm that Fos activation in SUS mice was due to the presence of a social target, we performed an additional RI test following CSDS with a novel object versus a juvenile intruder. Under these conditions, we found that Fos activity was significantly higher in SUS mice following juvenile interaction compared with novel-object interaction. Although we observed a slight increase in Fos activity following both novel-object and juvenile interaction in RES mice, there were no significant differences in time spent between them (Extended Data Fig. 3j,k).

Fig. 2: NT^{LS} neuronal activity is associated with altered social behaviours in susceptible mice.



a, Timeline of iDISCO+ Fos analysis from the RI test after CSDS. Timed sac, mice were perfused 90 min after RI test. **b**, Mouse brain before and after iDISCO+ clearing. **c**, Autofluorescence and Fos signal from lightsheet imaging. **d**, ClearMap analysis showing differentially activated brain regions from RES versus SUS mice. Arrowheads indicate LS. **e**, Neurotensin expression from Allen Brain Atlas ISH data. **f**, Timeline of ISH. **g**,**h**, Multiplex ISH (**g**) showing Fos expression (**h**) in NT neurons in females (one-way ANOVA, F (2, 6) = 7.887, P = 0.0209, n = 3 mice per group, three slices per mouse) and males (F (2, 10) = 13.13, P = 0.0016, n = 3 (CTRL), 4 (RES), 6 (SUS), three slices per mouse); scale bars, 50 µm. LV, lateral ventricle. **i**, Timeline of slice electrophysiology (ePhys) following CSDS. **j**, $eYFP^+$ NT^{LS} neurons patched in whole-cell configuration. **k**, Current–frequency curve showing counts of action potentials evoked by incremental steps of injected current. NT^{LS} neurons from SUS mice (n = 55 neurons) compared with RES mice (two-way ANOVA, P < 0.0001, n = 19). **l**, Resting membrane potential (RMP) for SUS and RES mice (two-tailed Mann–Whitney test, P = 0.0336, n = 4 (RES), 9 (SUS). **m**, Correlation between SI ratio and firing rate evoked by a 100 pA step current (Pearson's correlation, R^2 = 0.34, P = 0.0351). Each dot represents the mean value per mouse for RES (red, n = 4) and SUS (black, n = 9) mice. **n**, Sample traces of excitability for RES (red) and SUS (black) mice following 100 pA current injection. *P < 0.05, **P < 0.01. All data expressed as mean ± s.e.m.

Source data

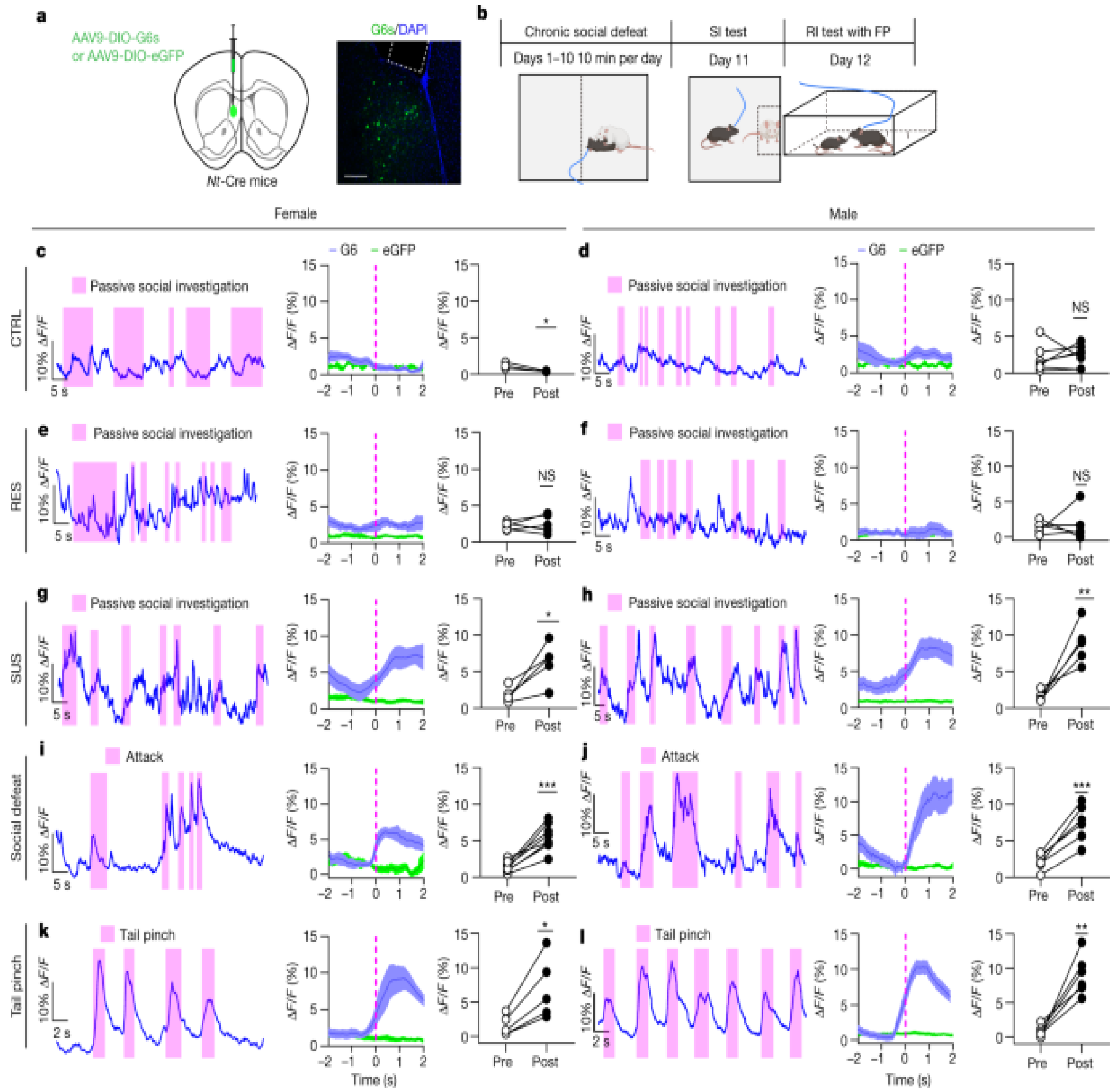
Due to its dense reciprocal interconnections throughout the brain's primary reward centres, the LS is often thought of as a nexus for mood[13,14,15,16], motivation[17,18] and spatial information processing[19]. Interestingly, in SUS mice following juvenile RI we found that most Fos-expressing neurons were located specifically in the lateral-ventral portion of the LS. Using the Allen Brain Atlas in situ hybridization (ISH) database[20], we found several genes expressed specifically in the lateral-ventral portion of the LS, including neurotensin (*Nt*) (Fig. 2e) and corticotrophin-releasing hormone receptor 2 (*Crhr2*). Oxytocin receptor (*Oxtr*) was expressed specifically in the lateral portion, somatostatin (*Sst*) was expressed mainly in the dorsal-lateral portion and dopamine receptor D3 (*Drd3*) was expressed only in the medial portion of the LS. Several recent studies have examined the role of these

molecularly defined cell types in regulation of behaviour, including *Sst*$^{+}$ neurons in fear conditioning[21], *vGAT*$^{+}$ and *Nt*$^{+}$ neurons in stress-suppressed feeding[22,23], *Crhr2*$^{+}$ neurons in anxiety-like behaviour[24], as well as *Oxtr*$^{+}$ and *Drd3*$^{+}$ neurons in social fear[25] and social dysfunction[26]. To determine which cell type was activated by juvenile social interaction in SUS mice, we performed multiplex ISH on brain slices from CSDS mice following juvenile RI (Fig. 2f). We found over 94% colocalization between *Nt* and *Fos* in SUS mice, with very limited expression of *Fos* in *Nt*$^{-}$ cells (Fig. 2g,h and Extended Data Fig. 4a). Around 100% of all *Nt*$^{+}$ cells were GABAergic (Extended Data Fig. 4b,c). Interestingly, *Nt* and *Crhr2* messenger RNA were largely colocalized in the anterior part but not in the posterior part of the LS, where we found significant increases in Fos levels following juvenile RI in SUS mice (Extended Data Fig. 4d,e). *Nt*$^{+}$ neurons had an overlap of about 5% with *Drd3*$^{+}$ and of about 20% with *Oxtr*$^{+}$ neurons (Extended Data Fig. 4f–i). Interestingly, we also found an increase in *Sst*$^{+}$ *Fos*$^{+}$ neurons in SUS mice following juvenile social interaction relative to CTRL mice, but not between RES and SUS mice (Extended Data Fig. 4j,k). Last, we found no differences in *Nt*$^{-}$ *Fos*$^{+}$ neurons between CTRL, RES and SUS mice (Extended Data Fig. 4l). Together these data highlight a potentially strong involvement of NT^{LS} neurons in social reward deficits in SUS mice.

To confirm that NT^{LS} neurons were indeed hyperactivated in SUS mice following interaction with a juvenile, we used a whole-cell slice electrophysiological approach to record NT^{LS} neurons in defeated male mice following a juvenile RI test (Fig. 2i,j). We found that NT^{LS} neurons from SUS mice showed increased excitability (Fig. 2k,n), as well as decreased resting membrane potential (Fig. 2l and Extended Data Fig. 5a), when compared with RES mice. Interestingly, we also found that the excitability of these cells was negatively correlated with the SI ratio observed following CSDS (Fig. 2m). These changes in intrinsic properties of NT^{LS} neurons suggest that CSDS induces lasting adaptations in these cells, which mediate social dysfunction. Interestingly, we found no differences in other properties of these cells (action potential threshold, amplitude, half-width or fast hyperpolarization; Extended Data Fig. 5a), confirming that CSDS specifically increases the excitability of these cells in SUS mice.

To further investigate NT^{LS} neuron activity in vivo during social encounters with juveniles, we injected Cre-dependent adeno-associated virus (AAV)-DIO-GCaMP6s into the LS of *Nt*-Cre transgenic mice to label NT^{LS} neurons (Fig. 3a). We then measured fluorescent Ca^{2+} activity by fibre photometry (FP) in CTRL, RES and SUS mice during juvenile RI (Fig. 3b and Extended Data Fig. 5b). We found no increase in NT^{LS} neuron activity in CTRL (Fig. 3c,d) and RES (Fig. 3e,f) mice in response to juvenile approach, but SUS mice exhibited significantly higher activity (Fig. 3g,h). Surprisingly, the magnitude (approximately 5–10% change in fluorescence ($\Delta F/F$)) of increase in NT^{LS} neuron activity during juvenile approach was similar to that observed when unstressed CTRL mice encountered an aversive experience, such as coming under attack by an aggressive CD-1 mouse (Fig. 3i,j) or experiencing a painful investigator-administered tail pinch (Fig. 3k,l). Moreover, NT^{LS} neuron activity showed no change following palatable food consumption (Extended Data Fig. 5c). These findings are consistent with the idea that NT^{LS} neurons respond to aversive, but not to rewarding, stimuli. We further tested NT^{LS} neuron activity during sCPP conditioning and found that NT^{LS} neurons in SUS mice showed higher activity during the juvenile-paired conditioning session, with no changes observed in CTRL or RES mice (Extended Data Fig. 5d). On the basis of these data, we suggest that, following CSDS, SUS mice may overgeneralize social threat cues and perceive juveniles as social threats, similar to that observed when being attacked by a highly aggressive CD-1 mouse.

Fig. 3: In vivo NT^{LS} activity in different social and stress contexts.



a, AAV-DIO-GCaMP6s injection and expression in LS. **b**, Timeline of FP experiments. **c**–**h**, Left, representative Ca^{2+} trace of NT^{LS} neurons during resident intruder test (pink strips indicate passive social investigation); middle, peri-event plot of NT^{LS} neuron activity 2 s before and after intruder approach; right, statistics for neuron activity 2 s before and 2 s after social events in CTRL females (paired two-tailed *t*-test, $t6 = 3.379$, $P = 0.0149$, $n = 7$) (**c**) and males ($t6 = 0.5081$, $P = 0.6295$, $n = 7$) (**d**); in RES females ($t4 = 0.6528$, $P = 0.5495$, $n = 5$) (**e**) and males ($t4 = 0.2939$, $P = 0.7834$, $n = 5$) (**f**); and in SUS females ($t4 = 3.772$, $P = 0.0196$, $n = 5$) (**g**) and males ($t4 = 4.844$, $P = 0.0084$, $n = 5$) (**h**). **i**–**l**, Left, representative Ca^{2+} trace of NT^{LS} neurons in CTRL mice during social defeat and tail pinches; middle, peri-event plot of

NT^{LS} neuron activity 2 s before and 2 s after attack/tail pinch; right, statistics of neuron activity 2 s before and 2 s after event in female defeat ($t7 = 6.852$, $P = 0.0002$, $n = 8$) (**i**), male defeat ($t6 = 6.973$, $P = 0.0010$, $n = 7$) (**j**), female tail pinch ($t4 = 3.988$, $P = 0.0163$, $n = 5$) (**k**) and male tail pinch ($t5 = 6.137$, $P = 0.0017$, $n = 6$) (**l**). All data were analysed using paired two-tailed *t*-test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. All data expressed as mean ± s.e.m. Scale bar, 100 μm. The illustrations in **b** were created with BioRender (https://biorender.com).

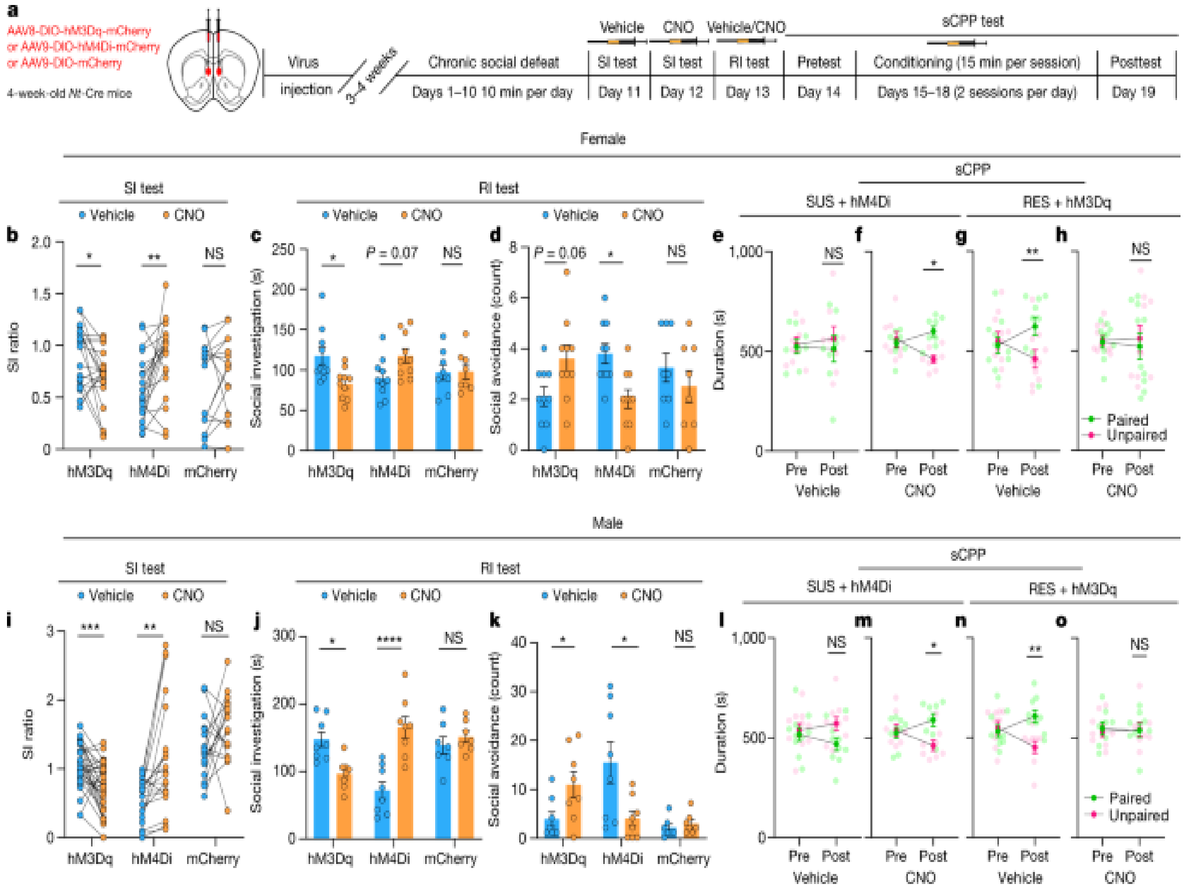
Source data

NT^{LS} neurons modulate social behaviour

To assess whether NT^{LS} neurons regulate social behaviours following CSDS, we utilized viral vectors expressing designer receptors exclusively activated by designer drugs (DREADDs) to bidirectionally manipulate the activity of NT^{LS} neurons during SI with a CD-1 mouse, and also during juvenile RI and sCPP. About 3–4 weeks before CSDS, we injected AAV-DIO-hM3Dq, AAV-DIO-hM4Di or AAV-DIO-mCherry viruses into the LS of 4-week-old *Nt*-Cre mice (Fig. 4a and Extended Data Fig. 6a). Mice were randomly assigned to CTRL or CSDS conditions. For inhibition studies with hM4Di to show necessity, we used only SUS mice whereas for activation studies aimed at showing sufficiency we used only RES mice (note: different baseline SI for vehicle-treated SUS hM4Di versus RES hM3Dq mice in Fig. 4). Testing was performed using a within-subjects design in which mice were first tested for SI 30 min after vehicle injection; then, 30 min before the second SI, mice were injected intraperitoneally with clozapine N-oxide (CNO). We found bidirectional effects of NT^{LS} neuron modulation on SI in both females and males, with increased activity reducing SI in RES mice and decreased activity enhancing SI in SUS mice (Fig. 4b,i). Micewere then split into two groups for RI, ensuring that the SI ratio was balanced across groups; mice received either vehicle or CNO during the RI test. Inhibition of NT^{LS} neurons increased social investigation time and normalized avoidance behaviour in both sexes (Fig. 4c,d,j,k). For sCPP, we treated hM4Di-injected SUS mice and hM3Dq-injected RES mice with vehicle or CNO during social conditioning sessions. We found that, by inhibiting NT^{LS} neurons in

SUS mice, we could normalize preference for the social conditioned compartment to CTRL or RES levels in both sexes (Fig. 4e,f,l,m). Conversely, by activating NT^{LS} neurons in RES mice, we were able to reduce social investigation and social preference compared with their vehicle-treated controls in both sexes (Fig. 4g,h,n,o). Thus, we find that activation of NT^{LS} neurons resulting from social trauma is both necessary and sufficient to elicit social behaviour deficits. Interestingly, activation of NT^{LS} neurons in stress-naïve mice affected neither SI ratio nor sCPP (Extended Data Fig. 6b–d), which suggests that a history of social trauma is critical. In line with this, we find no effect of non-social stressors, such as CVS, on social reward (Extended Data Fig. 6e,f), despite the fact that both CSDS and CVS similarly reduce preferences for natural rewards like sucrose[27]. Consistent with this, a recent study showed that ventral CA3 neurons projecting to the LS play a role in acute social stress-induced avoidance[28], but not in unstressed mice[29]. To test whether NT^{LS} neurons can more generally regulate reward or aversion behaviour, we utilized a real-time place preference assay (RTPP) in stress-naïve mice and found no effect of optogenetic stimulation of NT^{LS} neurons on preference (Extended Data Fig. 6g,h). Taken together, these data support the idea that NT^{LS} circuits modulate social behaviours in a context-dependent fashion.

Fig. 4: Chemogenetic manipulations of NT^{LS} neurons alter social behaviour following CSDS.



a, Experimental timeline for DREADD experiments. **b**–**d**,**i**–**k**, SI ratio, social investigation and social avoidance following chemogenetic activation (RES mice) or inhibition (SUS mice) of NT^{LS} neurons during social test following CSDS (two-way repeated-measures ANOVA, female: F (2, 53) = 9.785, P = 0.0002, n = 18 (hM3Dq), 20 (hM4Di), 16 (mCherry) (**b**), F (2, 25) = 5.807, P = 0.0085 (**c**), F (2, 25) = 5.906, P = 0.0079, n = 9 (hM3Dq), 10 (hM4Di), 8 (mCherry) (**d**); male: F (2, 64) = 12.96, P < 0.0001, n = 30 (hM3Dq), 20 (hM4Di), 17 (mCherry) (**i**), F (2, 20) = 19.46, P < 0.0001 (**j**), F (2, 20) = 10.12, P = 0.0009, n = 8 (hM3Dq), 8 (hM4Di), 7 (mCherry) (**k**)). **e**–**h**,**l**–**o**, Social preference rescued by inhibition of NT^{LS} neurons in female SUS mice (two-way repeated-measures ANOVA, CNO (**e**), F (1, 14) = 7.272, P = 0.0174, n = 8; vehicle (**f**), F (1, 14) = 0.3070, P = 0.5883, n = 8); and in male SUS mice (CNO (**l**), F (1, 14) = 4.710, P = 0.0477, n = 8; vehicle (**m**), F (1, 18) = 1.627, P = 0.2183, n = 10). Activation of NT^{LS} populations in RES females (two-way repeated-measures ANOVA, CNO (**g**), F (1, 18) = 0.1653, P = 0.6891, n = 10; vehicle (**h**), F (1, 18) = 8.490, P = 0.0093, n = 10); and in RES males (CNO (**n**), F (1, 14) = 0.2221, P = 0.6447, n = 8;

vehicle (**o**), $F(1, 16) = 9.283$, $P = 0.0077$, $n = 9$) blocked social CPP formation. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. All data expressed as mean ± s.e.m.

Source data

To determine whether NT^{LS} neurons encode context-specific information related to non-social stressors, we exposed WT mice to chronic restraint stress (CRS) and then performed an interaction test with a new restraint tube. We found that mice exposed to CRS had a longer latency to approach the tube and reduced time spent investigating the tube (Extended Data Fig. 7a,b). We then silenced NT^{LS} neurons with an inhibitory DREADD and found that this partially rescued tube avoidance (Extended Data Fig. 7c,d). In a separate group, we paired WT mice with juvenile bedding/odour during CRS (CRSO) and found no differences in the juvenile RI test, suggesting that mice do not generalize avoidance to a juvenile social target based on exposure to these olfactory cues (Extended Data Fig. 7e,f). Overall, these data suggest that NT^{LS} neurons are involved in more general computations that use past information from stressful or threatening situations to guide future behaviours towards cues associated with the same or similarly threatening/stressful situations. Last, we found a role for NT^{LS} neurons in mediating anxiety-related behaviours, such as the elevated plus maze (EPM), marble burying test and open field test (OFT) (Extended Data Fig. 7g–j). Together these results highlight the LS as a critical node in the regulation of emotional behaviours, particularly in response to aversive/stressful experiences.

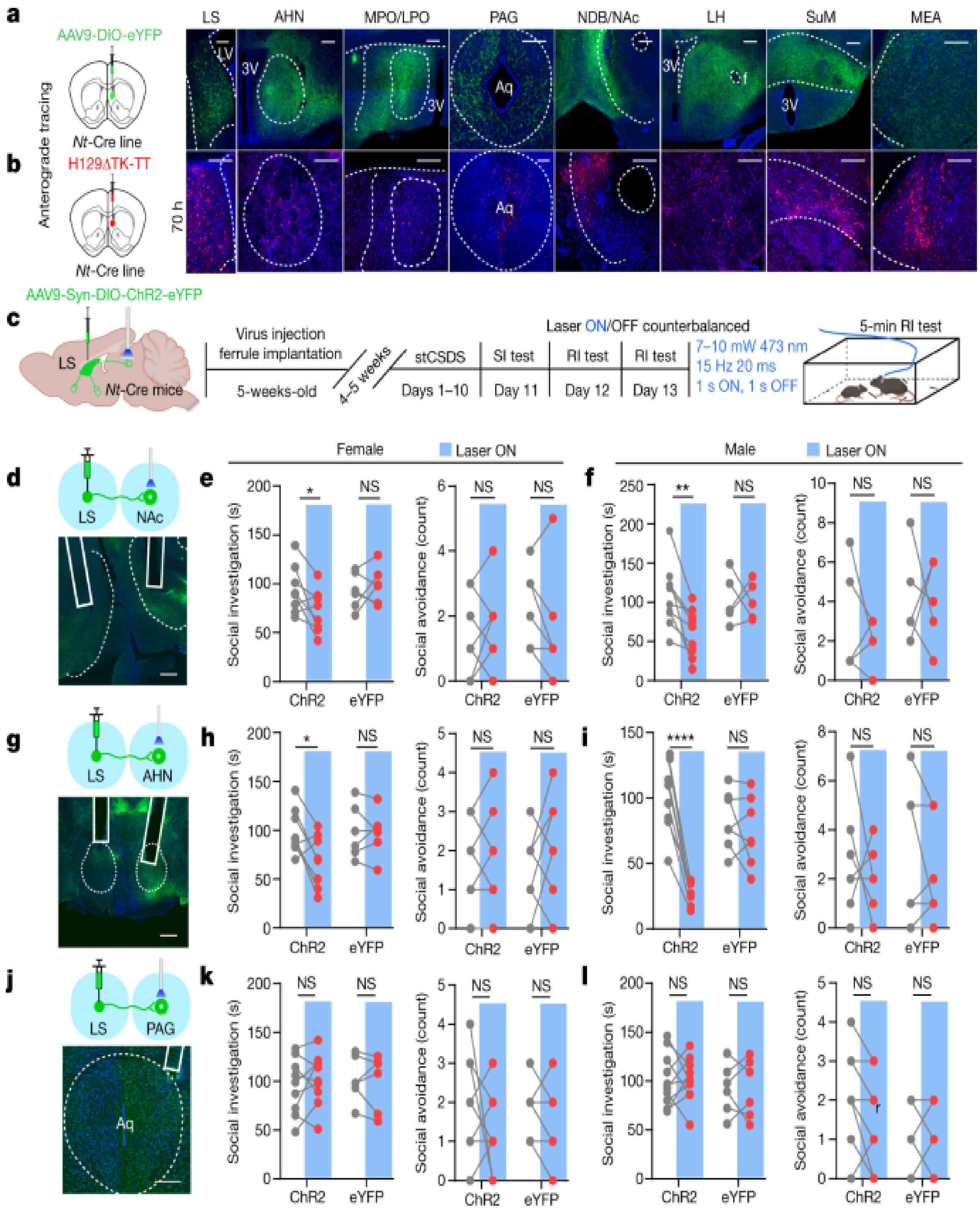
NT^{LS} circuitry regulates social behaviour

The LS contains long-range GABAergic projection neurons[30] and has topographically distributed, wide-range input–output projections[18,31,32]. To determine the output patterns of NT^{LS} neurons, we applied multiple viral-mediated anterograde tracing tools. First, we injected AAV-DIO-enhanced yellow fluorescent protein (eYFP) into the LS of *Nt*-Cre mice and imaged $eYFP^+$ axon terminals throughout the brain (Fig. 5a). We then used HSV-1 (H129ΔTK-TT) for anterograde trans-synaptic tracing[33] to verify which

regions form monosynaptic connections with NT^{LS} neurons (Fig. 5b and Extended Data Fig. 8a). Interestingly, many of the downstream regions identified, such as the medial-lateral preoptic area (LPO/MPO), nucleus accumbens (NAc), anterior hypothalamic nucleus (AHN), lateral hypothalamus (LH), periaqueductal grey (PAG), medial amygdala (MEA) and supramammillary nucleus (SuM), are all involved in various aspects of social behaviour[34] or conditioned reward[35]. Among these regions, the NAc is involved in social reward[36,37] and stress susceptibility[35,38], and the PAG in social aggression[39], as well as in defensive and escape behaviours[40,41]. Although the AHN plays a role in defensive behaviour[42] and parental behaviour[43], its role in social reward remains unknown. We wanted first to determine whether the same or different NT^{LS} neurons project to each of these sites. We injected a Cre-dependent retrograde AAV (rgAAV-DIO) expressing tdTomato into the AHN, NAc or PAG of *Nt*-Cre mice (Extended Data Fig. 8b). In the same mice, rgAAV-DIO-eYFP was injected into the alternate regions and we visualized overlap between tdTomato and eYFP in NT^{LS} neurons. We also injected cholera toxin subunit B (CTB) into the NAc (CTB488), AHN (CTB555) and PAG (CTB647) (Extended Data Fig. 8d) and found similar results: AHN/NAc-, AHN/PAG- and NAc/PAG-projecting LS neurons showed little overlap (Extended Data Fig. 8c,e), further confirming that LS neurons projecting to these regions represent mostly separate subpopulations. To investigate the function of these NT^{LS} circuits, we injected AAV-DIO-ChR2(H134R) into the LS of 5-week-old *Nt*-Cre mice and implanted ferrules in the NAc, AHN or PAG. Three weeks later, mice underwent a subthreshold CSDS (stCSDS) and social behaviour was tested during a 2-day, 5 min juvenile RI test in which laser on/off order was counterbalanced (Fig. 5c). Activation of NT^{LS}→AHN or NT^{LS}→NAc circuits decreased active social investigation time without affecting social avoidance behaviour during passive social bouts initiated by the juvenile (Fig. 5d–i). However, activation of NT^{LS}→PAG circuit had no effect on either social investigation time or social avoidance (Fig. 5j–l). To further validate whether manipulation of NT^{LS}→AHN or NT^{LS}→NAc circuits can bidirectionally modulate social interaction, we injected AAV-DIO-eNpHR3.0 into the LS and then performed CSDS (Extended Data Fig. 9a). We found that inhibition of the NT^{LS}→AHN or NT^{LS}→NAc circuits increased social investigation and partially decreased social avoidance

during the RI test (Extended Data Fig. 9b–e). To test whether these pathways bidirectionally control social preference, we injected either AAV-DIO-ChR2 or AAV-DIO-eNpHR3.0, exposed mice to social defeat stress and then performed optical stimulation of NT^{LS}→AHN or NT^{LS}→NAc circuits during the social conditioning session. As expected, we found that bidirectional regulation of both pathways affected sCPP (Extended Data Fig. 9f–m). The eNpHR3.0 manipulation seemed to have more subtle effects in general, possibly due to its poor efficacy in presynaptic inhibition. Recently developed G-protein-coupled optogenetic tools[44,45] may provide a more convincing method for long-range presynaptic inhibition in future studies.

Fig. 5: Optogenetic manipulation of NT^{LS} downstream regions alters social behaviour in CSDS mice.



a, Anterograde AAV-DIO-eYFP tracing from NT^{LS} neurons. **b**, Anterograde HSV-1 (H129ΔTK-TT) trans-synaptic tracing (70 h post injection) verifies monosynaptic connections between NT^{LS} neurons and regions shown in **a**. **c**, AAV-DIO-ChR2 injection and timeline for resident intruder optogenetics experiments. **d**–**l**, ChR2 axon terminal activation in NAc (**d**), AHN (**g**) and

PAG (**j**) during RI test in females (NAc (**e**), social investigation, F (1, 12) = 4.836, P = 0.0482; social avoidance, F (1, 12) = 2.935, P = 0.1123, n = 8 (ChR2), 6 (eYFP); AHN (**h**), social investigation, F (1, 12) = 4.947, P = 0.0461, social avoidance, F (1, 12) = 0.8571, P = 0.3728, n = 8 (ChR2), 7 (eYFP)); PAG (**k**), social investigation, F (1, 13) = 0.6986, P = 0.4183; social avoidance, F (1, 13) = 0.07324, P = 0.7909, n = 8 (ChR2), 6 (eYFP); and in males (social investigation, NAc (**f**), F (1, 13) = 4.540, P = 0.0528; social avoidance, F (1, 13) = 0.2848, P = 0.6026, n = 9 (ChR2), 5 (eYFP); AHN (**i**), social investigation, F (1, 13) = 28.94, P = 0.0001, social avoidance, F (1, 13) = 0.06521, P = 0.8024, n = 8 (ChR2), 7 (eYFP); PAG (**l**), social investigation, F (1, 14) = 0.002038, P = 0.9646; social avoidance, F (1, 14) = 1.750, P = 0.2071, n = 9 (ChR2), 6 (eYFP) (**f**)). Two-way repeated-measures ANOVA was performed for all comparisons. *P < 0.05, **P < 0.01, ****P < 0.0001. All data expressed as mean ± s.e.m. Scale bars, 200 μm. The illustrations in **c** were created with BioRender (https://biorender.com).

Source data

To confirm that these projections were monosynaptic and inhibitory, we injected AAV-DIO-ChR2 into the LS of *Nt*-Cre mice and performed ex vivo whole-cell electrophysiology with ChR2-assisted circuit mapping of NT^{LS}→NAc and NT^{LS}→AHN pathways. Our data show both pathways to be monosynaptic (with TTX), inhibitory (Cs-based internal, clamped at 0 mV) and GABAa-dependent (SR-95531, Gabazine) (Extended Data Fig. 10a,b). We also validated that 15 Hz of blue-light stimulation can reliably evoke NT^{LS} neurons (Extended Data Fig. 10c). Because it has been reported that other cell types in the LS can modulate stress behaviours under different conditions, we tested whether non-NT neurons in the LS also play a role in social trauma-induced social deficits by injecting AAV-Flpo and AAV-CreOff-FlpOn-ChR2 viruses into the LS of *Nt*-Cre mice to label non-NT neurons with ChR2 (Extended Data Fig. 10d). We first validated the specificity of this approach using Multiplex ISH (Extended Data Fig. 10e,f), and found very little overlap between ChR2 and NT. We next validated stimulation parameters for ChR2 using slice electrophysiology and confirmed that 15 Hz reliably activated non-NT neurons in the LS (Extended Data Fig. 10g). We then stimulated non-NT neurons in the LS in vivo at 15 Hz during the RI test following CSDS and found no effect on social

interaction (Extended Data Fig. 10h). Taken together, these results suggest that the activation of inhibitory NT^{LS} projections to the AHN and NAc is both necessary and sufficient to alter social investigation and social preference of mice following traumatic social experience.

Discussion

Many components of social behaviour, including its rewarding properties, are evolutionarily conserved between humans and rodents[46,47]. Although it is well established that social stress leads to the development of depression, anxiety[48] and post-traumatic stress disorder[38], the neural circuits that mediate the negative consequences of social stress—particularly with regard to social reward—are not well defined. We view preclinical social stress models as imperative to this early-phase work so that we can define potential circuit mechanisms of trauma-impaired social reward to inform future studies in humans.

Utilizing an unbiased approach, we identified the LS as one of the most highly regulated regions activated in both male and female SUS mice in response to a normally rewarding social target, suggesting that it might be a particularly important region in regard to explaining the common social deficits exhibited by both sexes. Detailed analysis of the LS identified a specific population of GABAergic projection neurons expressing the neuropeptide neurotensin. In unstressed mice we found these cells to be responsive during situations of threat, including in response to aggressive attack behaviour. However, following chronic social trauma in SUS mice we found that these neurons generalize their responses to non-threatening social situations, including during interactions with non-aggressive juvenile mice. Notably, NT^{LS} and *Drd3*$^{+}$ neurons exert opposing functions to control social behaviour following stress (Fig. 4 and ref. [26]). *Drd3*$^{+}$ and *Nt*$^{+}$ neurons are topographically distinct in the LS and it is likely that they have different input/output projection patterns, and possibly even form distinct synaptic connections within the LS. Thus, we hypothesized that NT^{LS} neurons might play a unique role in regulation of social reward by inhibiting downstream reward centres. Indeed, anterograde tracing studies identified known reward centres—including the NAc and AHN—as receiving moderate/dense

innervation from NT^{LS} neurons, and activation of these inputs reduced social interaction and context-dependent social reward with a juvenile.

Because anxiety is well known to inhibit adaptive social behaviours; one critical question is whether NT^{LS} neurons are encoding social aversion or whether they simply encode a generalized state of anxiety that impairs social behaviours. According to our data, generalized anxiety states measured by EPM/OFT are separable from social behaviour deficits: (1) when we stimulate NT^{LS} neurons in social stress-naïve mice, we are able to produce a generalized exploratory deficit in the EPM/OFT (Extended Data Fig. 7); however, such stimulation does not induce avoidance of a social target (Extended Data Fig. 6b). (2) Both RES and SUS mice in the CSDS model exhibited anxiety-like behaviours in the OFT and EPM, yet only SUS mice exhibited social avoidance and reduced social reward. (3) Although CVS produces an increase in generalized anxiety-like behaviour, it has no effect on social interaction or social reward (Extended Data Fig. 6e,f). Thus, in addition to regulation of generalized anxiety states, NT^{LS} neurons encode contextual information about stressful/traumatic past experiences to guide future behavioural responses.

Overall our findings demonstrate that, in both male and female SUS mice, rewarding social targets are perceived as stressful or threatening, which engages NT^{LS} circuitry and impairs social reward processing in a context-dependent manner. Interestingly, in studies of patients with depression and comorbid social anxiety disorder, it was shown that social trauma abnormally increases the representation of social threat[49]. Furthermore, children who have experienced trauma are reported to exhibit heightened perceptual threat sensitivity, negative and neutral emotion misclassification and attention biases towards threat-related cues[50]. Our research thus provides an important foundation for understanding the neural mechanisms underlying post-trauma social reward processing. Future studies in humans to test the relevance of LS circuitry in mediation of social threat perception and reward sensitivity in victims of trauma will be highly informative.

Methods

Animals

Wild-type C57BL/6J mice, 7–8 weeks old (males, 22–26 g; females, 18–22 g; Jackson Laboratory) were used as experimental mice in CSDS studies; 4–6-week-old C57BL/6J mice (Jackson Laboratory) were used as new intruders in both the RI test and sCPP test; 16–24-month-old male CD-1 (ICR) mice (sexually experienced retired breeders; Charles River Laboratories) were used as aggressors for male CSDS. *ERα*-Cre mice (017911, B6N.129S6(Cg)-Esr1$^{tm1.1(cre)And}$/J; Jackson laboratory) were crossed with CD-1 females to obtain F1 males, which were used as aggressors for female CSDS. *Nt*-Cre (01752, B6;129-Nts$^{tm1(cre)\ Mgmj}$/J; Jackson Laboratory) homozygous mice were crossed with WT C57BL/6 J mice, and the F1 generation was used as experimental mice in the CSDS studies. Littermates were randomly assigned to experimental groups. All mice were allowed 1 week of acclimation to the housing facilities before the start of experiments. WT CD-1 and F1 *ERα*-Cre mice were single housed, *Nt*-Cre and WT C57BL/6J mice were housed in groups of between three and five. All mice were maintained on a 12/12-h light/dark cycle (07:00–19:00) with ad libitum access to food and water. Housing and experimental rooms were maintained at 20–22 °C and 40–60% humidity. Experiments were conducted during the light phase. Procedures were performed in accordance with the National Institutes of Health Guide for Care and approved by the Use of Laboratory Animals and the Icahn School of Medicine at Mount Sinai Institutional Animal Care and Use Committee. Additional information about mice used in this study can be found in the Life Sciences Reporting Summary.

Aggressor screening, CSDS and stCSDS

Female[11] and male[10] aggressor screening for CSDS and SI tests was performed as previously described. Experimental males were single housed after CSDS, and females were group housed during CSDS but single housed after defeat. Defeat was halted if an intruder showed any signs of injury. An all-male CSDS lasted 10 min per day for 10 days an all-female CSDS lasted 5 min per day for 10 days; stCSDS lasted for 5 and 2 min per day for males and females, respectively, for 10 days.

Chronic variable stress

CVS was modified from our previous work[51]. Male and female mice were randomly assigned to CTRL and CVS groups. CVS groups underwent 28 days of stress with one stressor per day, the stressors consisting of 1 h foot shock (random shock 60 times in 1 h), 1 h tail suspension and 1 h restraint.

Chronic restraint stress

Male mice were randomly assigned to CTRL and CRS groups. The CRS group underwent 28 days of 1 h restraint stress each day. For the juvenile odour-paired CRS, mice were restrained in a 50 mL restrainer and put in a new cage with bedding from a same-sex C57BL/6J juvenile mouse.

Social interaction test

SI tests were performed 24 h after the last defeat, as described previously[10]. Mice were habituated in the testing rooms for 1 h before testing and all testing was performed under red-light conditions. SI tests were performed with mice freely exploring in a target-free arena (44 cm (w) × 44 cm (d) × 38 cm (h)) for 2.5 min, followed by another 2.5 min target-present (CD-1 or *ERα*-Cre mice) session during which target mice were confined in a wire-mesh enclosure (10 cm (w) × 6.5 cm (d) × 38 cm (h)). The 'interaction zone' of the test arena encompassed a 14 cm × 24 cm rectangular area projecting 8 cm around the wire-mesh enclosure. The 'corner zones' encompassed a 9 cm × 9 cm area projecting from both corner joints opposing the wire-mesh enclosure. We calculated SI ratio as the ratio of time spent in the interaction zone with a CD-1 or F1 *ERα*-Cre mouse present over time spent with the target absent. All mice with a SI ratio over 1 were classified as RES mice and all with a ratio below 1 as SUS mice. Corner ratio was calculated as the ratio of time spent in the corner zone with an adult CD-1 or F1 *ERα*-Cre target mouse present over time spent when the target mouse was absent.

Resident intruder test

The RI test was modified from a previously described protocol[52]. After defeat and SI, mice were habituated in the testing rooms for 1 h before

testing, and all testing was performed under dim light conditions. Experimental mice were kept in their home cage, placed under an Ethovision camera, habituated for 2–3 min with their wired cage top removed and then intruders (mice or objects) were introduced into their home cage and allowed to interact freely for 5 min (RI test for iDISCO+ cohort lasted for 10 min, to maximumly stimulate Fos expression). Social investigation included the amount of active interaction including approach, close following and sniffing. Social avoidance was defined as the escaping from a juvenile mouse of the experimental mouse when approached and investigated by the former. A speed of more than 20cm s^{-1} was considered an escape. SUS mice typically escaped at speeds of 20–65 cm s^{-1} immediately to avoid social encounters following juveniles' approach/investigation. All experimental mice showing aggressive behaviours towards juveniles (around 1%) were excluded from analyses. All RI behaviours were scored blindly by experimenters.

Social conditioning place preference

The sCPP protocol, as published previously[53], consisted of three phases: pretest, social conditioning and posttest. Mice were habituated in the testing rooms for 1 h before conditioning or testing. All phases were conducted under red-light and sound-attenuated conditions. The CPP apparatus (Med Associates) has a neutral middle zone that allows for unbiased entry and two conditioning chambers with different walls and floors. On the pretest day, mice were introduced into the middle chamber and allowed to explore freely in all three chambers of the CPP box for 20 min. No group differences in bias for either chamber were found, and conditioning groups were balanced in an unbiased fashion to account for start side preference, as described previously[54]. The conditioning phase consisted of four consecutive days with two conditioning sessions each day: during the morning paired sessions (08:00–12:00), experimental mice were confined to the assigned chamber for 15 min with a new same-sex juvenile C57BL/6J intruder; during the afternoon unpaired session (13:00–17:00) mice were put into the opposite chamber without a social target for 15 min. For female sCPP, during conditioning the juvenile mice were confined in a wire-mesh cup, which we found was necessary for females to form CPP, whereas males formed a preference only when they were able to freely interact with the juvenile

outside the cup. All groups were counterbalanced for conditioning chamber. On posttest day, experimental mice were placed in the middle chamber of the CPP apparatus and allowed to freely explore all chambers for 20 min. Duration spent within either context was used to measure CPP. For chemogenetics experiments, CNO was administered during the full conditioning sessions. Behavioural analysis of sCPP data was performed by assessing (1) subtracted CPP (posttest phase duration spent in the intruder-paired chamber minus test phase duration spent in the intruder-unpaired chamber, accounting for test session behaviour only); and (2) group and individual durations in both pretest and posttest sessions.

Novel-object recognition and object location test

Novel-object recognition (NOR) and object location tests were performed as previously described[55]. Male mice were habituated in the testing room for 1 h before testing and then placed in the middle of an empty plexiglass open field (45 cm (w) × 45 cm (d) × 38 cm (h)) under dim light for 10 min (habituation phase). Twenty minutes after the habituation phase, mice were placed in the same open field with two objects (A and B) and allowed to explore for 10 min. Mice were then placed back in their home cage for 20 min before being put back into the open field with object B replaced by a new object, C. Mice were allowed to explore for 10 min. Following the NOR test, mice were transferred back to their home cage for 20 min before being returned to the open field, in which the location of object A was changed and the time spent interacting was recorded. Time spent with the new versus familiar object or location was recorded and scored by Ethovision software.

Elevated plus maze

The EPM was performed as previously described[11]. Male mice were habituated in the testing room for 1 h before testing and then placed in the middle of the plexiglass EPM under red light for 5 min. Each arm of the maze measured 12 × 50 cm^2. Behaviour was tracked using Noldus Ethovision (Noldus Interactive technologies). Total time spent in the open and closed arms was measured.

Open field test and locomotor measures

Open field test was performed as previously described[11]. Male mice were habituated in the testing room for 1 h before testing and then placed in the middle of the plexiglass arenas (44 × 44 × 35 cm^3) under red light for 10 min. Behaviour was tracked using Noldus Ethovision (Noldus Interactive technologies) to record the total distance moved, as well as the time spent in the centre (22 × 22 cm^2) versus outer zones.

Marble burying test

The marble-burying test was performed as previously described[56]. Male mice were habituated in the testing room for 1 h before testing. Fresh, unscented mouse corncob bedding (depth 5 cm) was put in standard rat cages (26 cm (w) × 48 cm (d) × 20 cm (h)) with filter-top covers, and another cage was inserted onto the surface of the bedding to create parallel lines on the bedding surface that could be used for marble placement. Standard glass toy marbles (1.6 cm diameter) were placed gently on the surface of the bedding in five rows of four. Marbles were washed in 70% ethanol, rinsed in distilled water and dried before each use. Mice were introduced into the corner of the cage to explore for 30 min with the filter-top covered on the cage. A marble was scoredas buried if two-thirds of its surface area was covered by bedding. A 2-day, DREADD-manipulated marble-burying test was performed using a within-subjects design; mice were given either CNO or vehicle in a counterbalanced way, and thus they received CNO or vehicle on the first day and the alternative on the second day.

Real-time place preference

The RTPP experiments was performed as previously described[54]: mice were placed in the centre of an arena (44 cm (w) × 44 cm (d) × 35 cm (h)) with a central divider and allowed to explore freely for 20 min. The time spent on each side was recorded using Noldus Ethovision (Noldus Interactive technologies). For the first 10 min of the test, one side of the open field was paired with 20 ms pulses of 15 Hz blue-light stimulation (473 nm, 7–10 mW, 1 s on, 1 s off). For the second 10 min of the test, laser stimulation was paired with the opposite side of the arena; this was done to minimize inherent bias toward one side of the arena. There was a 1-min interval

between the two phases. Total time spent in the unstimulated and stimulated sides was calculated and analysed.

Perfusion and brain tissue processing

For immunohistochemistry and iDISCO+, mice were euthanized by injection of 10% chloral hydrate and perfused transcardially with ice-cold 1× PBS (pH 7.4), followed by fixation with cold 4% paraformaldehyde in 1× PBS. Brains were postfixed for 12 h in the same fixative at 4 °C. For immunohistochemistry, coronal sections were prepared on a vibratome (Leica) at 50 µm to assess viral placement and for immunohistochemistry. For ISH, mouse brains were rapidly removed and flash-frozen in −30 °C isopentane for 5–10 s then kept at −80 °C until sectioning at 15 µm using a cryostat (Leica). Animals injected with AAV viruses were perfused at least 4 weeks after injection; animals injected with H129ΔTK-TT were perfused 48 and 70 h after injection.

IHC, ISH and confocal microscopy

For Fos IHC, slices were incubated for 2 h in blocking solution (3% normal donkey serum, 0.3% Triton X-100 in PBS) then incubated overnight in primary antibody (mouse anti-Fos, 1:1,000 (Santa Cruz Biotechnology, C-10)) at 4 °C. Slices were then washed in PBS for 3 × 20 min and incubated in secondary antibody (Cy2 (no. 711-225-152), Cy3 (no. 711-165-152), Cy5 (no. 711-175-152), AffiniPure Donkey Anti-Rabbit IgG (H+L), 1:1,000 (Jackson ImmunoResearch)) for 2 h at room temperature, then washed in PBS for 3 × 20 min before staining with DAPI (1 µg mL^{-1}, Sigma) for 20 min. Sections were then mounted with Eco-Mount (Life sciences) and coverslipped (Fisher). For Fos analysis, magnification of ×20 and tile-scan function were used to acquire the entire region of interest. Analysis of Fos-positive cells was performed using Fiji (NIH)[57]. For representative images of viral infection, images were acquired at ×10 magnification using the tile-scan function. For ISH, RNAscope Multiplex Fluorescent Kits (Advanced Cell Diagnostics) were used according to the manufacturer's instructions. Briefly, fresh-frozen brains were slide mounted at 15 µm thickness, fixed for 15 min in cold 4% PFA and dehydrated serially with 50, 70 and 100% EtOH/H_2O for 2 min each, followed by 20 min Protease IV (RNAscope)

treatment. Proprietary probes (Advanced Cell Diagnostics) for Fos (316921, accession no. NM_010234.2); *Sst* (404631-C2, accession no. NM_009215.1), *Gad67* (400951-C2, accession no. NM_008077.4), *Oxtr* (412171-C2, accession no. NM_001081147.1), *Drd3* (447721-C, accession no. NM_007877.1) or *Crhr2* (413201-C2, accession no. NM_009953.3); *Nt* (420441-C3, accession no. NM_024435.2) were hybridized at 40 °C for 2 h then subjected to a series of amplification steps at 40 °C (1-FL, 30 min; 2-FL, 15 min; 3-FL, 30 min; 4-FL, 15 min). Reagent Alt-A was used for the fourth amplification step, with Channel 1 at 488 nm, Channel 2 at 550 nm and Channel 3 at 647 nm. Finally, slides were treated for 1 min with DAPI and immediately coverslipped with Eco-Mount. All slices were imaged using a Zeiss LSM 780 confocal microscope. Cells and Fos from all ISH and IHC images were counted blindly across groups.

iDISCO+ staining, imaging and ClearMap analysis

The iDISCO+ staining protocol was modified from http://www.idisco.info. Fixed whole brains were incubated with the primary Fos antibody (no. 226003, 1:1,000, Synaptic Systems) and secondary donkey anti-rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 (no. A-31573, 1:1,000, Thermo Fisher Scientific) for 7 days each. A LaVision lightsheet microscope with zoom body was used for half-brain sagittal scanning, with dynamic focus and a step size of 4 um. Cleared brains were processed as previously described using ClearMap[12]. Fos^+ cells were quantified using the cell detection module, with cell detection parameters optimized and validated based on the intensity and shape parameters of the signal. The autofluorescence channel was aligned to the Allen Institute's Common Coordinate Framework using the Elastix toolbox. Brain areas were collapsed into their parent region (for example, the rostroventral, caudodorsal and ventral lateral septum were combined into the 'lateral septal nucleus'). These decisions were made before analysis. To compare cell counts in RES and SUS animals, a negative binomial regression was applied using the glm.nb function from the MASS package in R. Group classifications were dummy coded (0 for the SUS group and 1 for the RES group). The maximum-likelihood coefficients α and β were determined through iterative reweighted least squares. A significant β means that group status is related to cell count number at the specified region of interest. The

z-values in Extended Data Fig. 2i correspond to this β coefficient, normalized by its sample standard deviation. *P* values were corrected for multiple comparisons using the Benjamini–Hochberg procedure to decrease false discovery rate. *Q*-values below 0.05 were considered significant.

Stereotaxic surgery and viral gene transfer

Nt-Cre mice (4–5 weeks old) were anaesthetized by intraperitoneal injection with a mixture of ketamine HCl (100 mg kg^{-1}) and xylazine (10 mg kg^{-1}) and positioned on a stereotaxic instrument (David Kopf Instruments). In the LS (from bregma: AP +0.7 mm; ML ±0.4 mm; DV −3.0 mm), 0.5 μL of virus was bilaterally infused using 33-gauge Hamilton needles over 5 min, with needles left in place for 5 min after injection. For DREADD virus delivery, 0.5 µl of AAV8-hSyn-DIO-hM3D-mCherry (2.0×10^{12} vg mL^{-1}, no. 44361-AAV8, Addgene), AAV9-hSyn-DIO-hM4D-mCherry (2.0×10^{12} vg mL^{-1}, no. 44362-AAV9, Addgene) or AAV9-hSyn-DIO-mCherry (2.0×10^{12} vg mL^{-1}, no. 50459-AAV9, Addgene) was injected into the LS. For anterograde tracing, 0.5 μL of AAV9-hSyn-DIO-EYFP (2.0×10^{12} vg mL^{-1}, no. 50457-AAV9, Addgene) or 0.15 µl of H129ΔTK-TT (4.0×10^{9} vg mL^{-1}, Center for Neuroanatomy with Neurotropic Viruses) was injected unilaterally into the LS. For retrograde tracing of LS downstream regions, 0.5 μL of retrograde AAV-DIO-EGFP/tdTomato (2.0×10^{12} vg mL^{-1}, nos. 50457-AAVrg and 28306-AAVrg, Addgene) was injected into the medial part of the NAc (from bregma: AP +1.5 mm; ML ±0.5 mm; DV −4.4 mm), AHN (from bregma: AP −0.7 mm; ML ±0.5 mm; DV −5.0 mm) or PAG (from bregma: AP −4.2 mm; ML ±0.2 mm; DV −2.5 mm). For CTB injection, 0.5 μL of Alexa Fluor 488-conjugated Cholera Toxin Subunit B (1.0 mg mL^{-1}, no. C-34775, Thermo Fisher) was injected into the NAc (from bregma: AP +1.5 mm; ML ±0.5 mm; DV −4.4 mm), 0.5 μL of Alexa Fluor 555-conjugated Cholera Toxin Subunit B (1.0 mg mL^{-1}, no. C-34776, Thermo Fisher) was injected into the AHN (from bregma: AP −0.7 mm; ML ±0.5 mm; DV −5.0 mm) and 0.5 μL of Alexa Fluor 647-conjugated Cholera Toxin Subunit B (1.0 mg mL^{-1}, no. C-34778, Thermo Fisher) was injected into the PAG (from bregma: AP −4.2 mm; ML ±0.2 mm; DV −2.5 mm). For optogenetics, 0.5 μL of AAV9-EF1a-DIO-EYFP (3.0×10^{12} vg mL^{-1}, no. 27056-AAV9, Addgene), AAV9-Ef1a-DIO eNpHR3.0-EYFP (3.0×10^{12} vg

mL^{-1}, no. 26966-AAV9, Addgene) or AAV9-EF1a-DIO-ChR2-EYFP (3.0×10^{12} vg mL^{-1}, no. 20298-AAV9, Addgene) was injected into either the LS (cell body stimulation) or downstream regions (terminal stimulation). For CreOff virus injection, AAV-EF1a-Flpo (2.0×10^{12} vg mL^{-1}, no. 55637-AAV1, Addgene) and AAV-nEF-Coff/Fon-ChR2(ET/TC)-EYFP (2.0×10^{12} vg mL^{-1}, no. 137141-AAV8, Addgene) were mixed 1:1 and injected into the LS. All AAV injections were performed 3 weeks before perfusion or behavioural experiments. For aggressors used in female CSDS, we targeted the VMHvl of *ERα*-Cre F1 mice as described previously[11,58]. For FP, 0.5 μL of AAV9-CAG-FLEX-G6s/EGFP virus (2.0×10^{12} vg mL^{-1}, no. 100842-AAV9, 51502-AAV9 Addgene) was injected unilaterally into the LS. For optogenetic (ChR2) and FP experiments, cannulae (ChR2: MFC_200/240-0.22_3mm_MF1.25_FLT; FP: MFC_200/250-0.57_3mm_MF1.25_FLT) were implanted at the same time as viral delivery (for LS local, fibres were implanted 0.2 mm above the injection site). For optogenetic (ChR2 and eNpHR3.0) experiments on NT^{LS} terminal stimulation, cannulae (MFC_200/240-0.22_MF1.25_FLT, 5 mm for NAc/AHN, 3 mm for PAG) were implanted into the NAc (from bregma: AP +1.5 mm; ML ±1.5 mm; DV −4.4 mm, 15° angle), the AHN (from bregma: AP −0.7 mm; ML ±1.5 mm; DV −4.8 mm, 10° angle) or PAG (from bregma: AP −4.2 mm; ML ±0.2 mm; DV −2.3 mm). For secure fixture of the optic fibre, dental cement (Grip cement; Dentsply) was added to the skull and around the fibres.

DREADD manipulation

For *ERα*-Cre mice (used for female CSDS), CNO (1 mg kg^{-1}, Tocris) was given intraperitoneally 30 min before CSDS[11]. For OFT, EPM and the marble-burying, SI and RI tests, CNO was given 30 min before the test; for sCPP, CNO was given 30 min before each conditioning session.

Optogenetics manipulation

For blue (ChR2) and orange (eNpHR3.0) light stimulation, optical fibres (BFP(2)_200/220/900-0.22_4m_FCM-2xMF1.25, Doric Lenses) were connected to either a 473 nm blue laser diode (no. BCL-473-050-M, Crystal Laser) or a 589 nm orange laser diode (no. MGL-III-589-50mW, Opto

Engine LLC) using a patch cord with a FC/PC adaptor (no. MFP_200/240/900-0.22_4m_FC-MF1.25, Doric Lenses). A function generator (no. 33220A, Agilent Technologies) was used to generate 20 ms blue-light pulses at 15 Hz, 1 s on/1 s off for all ChR2 experiments. Constant orange light was used for eNpHR3.0 experiments during the 5 min resident intruder test. For sCPP studies, orange light was delivered in a 4 min on/1 min off pattern. The intensity of light delivered to the brain was 7–10 mW. These parameters are consistent with previously validated and published protocols[24]. For all optogenetics tests, experimental mice were habituated to patch cords for 2 days before testing in RI. For RI experiments, mice were tested over 2 days, counterbalanced under laser-on and -off conditions. For social CPP tests, light was provided during the social conditioning session. For the RTPP test, blue-light delivery was controlled by TTL from Noldus Ethovision (Noldus Interactive technologies).

Ex vivo electrophysiology

AAV9-hSyn-DIO-EYFP (0.5 ul, 2.0×10^{12} vg mL^{-1}, Addgene) was injected bilaterally into the LS of 4-week-old male *Nt*-Cre mice. Two to three weeks after injection, the mice underwent CSDS. Before slice preparation, all mice were exposed to a 4–6-week-old, same-sex juvenile intruder for 5 min. About 20 min after the RI test, mice were anaesthetized using isoflurane. The brain was rapidly extracted and coronal sections (250 µm) sliced using a Compresstome (no. VF-210-0Z, Precisionary Instruments) in cold (0–4 °C) sucrose-based artificial cerebrospinal fluid (SB-aCSF) containing (in mM): 87 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 4 $MgCl_2$, 23 $NaHCO_3$, 75 sucrose and 25 glucose. After 60 min at 32 °C for recovery, slices were maintained in oxygenated (95% CO_2/5% O_2) aCSF containing (in mM): 130 NaCl, 2.5 KCl, 1.2 NaH_2PO_4, 2.4 $CaCl_2$, 1.2 $MgCl_2$, 23 $NaHCO_3$ and 11 glucose at room temperature for the remainder of the day, and transferred to a recording chamber continuously perfused at 2–3 mL min^{-1} with oxygenated aCSF. Patch pipettes (4–7 MΩ) were pulled from thin-walled borosilicate glass using a micropipette puller (no. P-97, Sutter Instruments) and filled with a K gluconate (KGlu)-based intrapipette solution containing (in mM): 116 KGlu, 20 HEPES, 0.5 EGTA, 6 KCl, 2 NaCl, 4 ATP and 0.3 GTP and 2 mg mL^{-1} biocytin (pH adjusted to 7.2). Cells were visualized using an

upright microscope with an IR-DIC lens and illuminated with a white light source (Scientifica). A 470 nm LED (no. pE-300^{ultra}, Cooled) illumination through the microscope objective was used for visualization of eYFP^{+} cells (using a bandpass filter cube, Olympus). Excitability was measured in current-clamp mode by injection of incremental steps of current (0–100 pA, +10 pA at each step). For recording of optically evoked inhibitory postsynaptic currents (oIPSCs), AAV9-EF1a-DIO-ChR2-eYFP (0.5 μL, 3.0 × 10^{12} vg mL^{-1}, Addgene) was injected bilaterally into the LS of 4-week-old male *Nt*-Cre mice. At 5–8 weeks post injection, coronal brain slices of NAc/AHN were prepared as described above and NAc/AHN neurons were recorded in voltage-clamp mode using an internal solution containing (in mM): 120 Cs-methanesulfonate, 10 HEPES, 10 Na-phosphocreatine, 8 NaCl, 5 TEA-Cl, 4 Mg-ATP, 1 QX-314, 0.5 EGTA and 0.4 Na-GTP. NT^{LS} terminals were stimulated through the microscope x40 objective (15 Hz, 5 ms per pulse, 470 nm; no. pE-300^{ultra}, CoolLed). oIPSCs were recorded at 0 mV in the presence of tetrodotoxin (TTX, 1 μM, Tocris) to probe monosynaptic effects. oIPSCs were blocked by bath application of gabazine (no. SR-95531, 10 μM, Tocris) confirming the GABAergic nature of the synaptic contact. Whole-cell recordings were performed using a patch-clamp amplifier (Axoclamp 200B, Molecular Devices) connected to a Digidata 1550 LowNoise acquisition system (Molecular Devices). Signals were low-pass filtered (Bessel, 2 kHz) and collected at 10 kHz using the data acquisition software pClamp 11 (Molecular Devices). Electrophysiological recordings were extracted and analysed using Clampfit (Molecular Devices). All groups were counterbalanced by days after defeat. All recordings were performed blind to experimental condition.

Fibre photometry

Fibre photometry was performed according to the Neurophotometrics manual and published protocols[59]. A fibre-optic patch cord (no. MFP_200/240/900-0.48_3m_FC-MF1.25, Doric Lenses) was attached to the implanted cannula with cubic zirconia sleeves covered with dark-coloured, shrinkable tubing. The other end of the fibre-optic cable was coupled to a Neurophotometrics LED port. The open-source Bonsai programme was used to control the system; 470 and 415 nm LED lights were used for GCaMP6s signal and autofluorescence measurement. Light at the fibre tip ranged from

40 to 80 μW and was constant across trials over testing days. Simultaneous recording of 40 fps from both 470 and 415 nm channels was achieved phase to phase and visualized via Bonsai. Three weeks after virus injection and ferrule implantation, when mice were around 8 weeks old, they underwent CSDS and SI; they were then habituated to the patch cord for 2 days and Ca^{2+} fluorescence was recorded during the RI test, social CPP conditioning session, stress and food reward tests. Once connected to the apparatus, mice were allowed to rest and habituate for 3–5 min before starting. For the RI test, we recorded Ca^{2+} fluorescence during 2 min of baseline activity without an intruder, followed by 5 min of intruder exposure. The food reward was performed in an open field, and peanut butter cups were placed in the arena close to the corners. All food-biting events were scored manually. MATLAB custom-coding was used for analysis of signal. The 415 nm channel served as the control channel and was subtracted from the GCaMP6s channel to eliminate autofluorescence, bleaching and motion effects. Change in fluorescence ($\Delta F/F$) was calculated as the percentage of mean fluorescence signal of GCaMP6s signal. In general, these motion artefacts had very little effect on overall GCaMP6s signal. Behavioural data were aligned with fluorescence recording data by dividing behavioural video frames with GCaMP6s signal frames. For analysis of LS GCaMP6s activity during discrete behaviours in the RI test, average $\Delta F/F$ (%) in the 2 s before and after a discrete event (passive social investigation) were compared. A passive social investigation was determined to occur at the moment of the intruder-initiated passive social approach.

Statistical analysis

All statistical details can be found in the figure legends, including type of statistical analysis used, P values, n, what n represents, degrees of freedom and t or F values. All t-tests, one-way ANOVA and repeated two-way ANOVA were performed using GraphPad Prism software (GraphPad Software Inc.). One-way ANOVA analysis was followed by Tukey's multiple-comparisons test, and two-way repeated-measures ANOVA analysis was followed by Šídák's multiple-comparisons test. Statistically significant differences are indicated in each figure (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$). For detailed P values please see Source data. Analyses of Fos staining, ISH data and behavioural videos during the

RI test were performed blinded to experimental conditions. Sample sizes were chosen according to previous experiments. For Extended Data tables and the Supplementary table, *P* values were corrected for multiple comparisons using the Benjamini–Hochberg procedure to reduce false discovery rate. *Q*-values below 0.05 were considered significant for all Extended Data tables.

Statistics and reproducibility

Figure 2c and Extended Data Fig. 3i were repeated in three separate cohorts per sex, with all showing similar results. Figure 5a,b (right) was repeated in three seperate male cohorts ($n = 6$) and in one female cohort ($n = 2$), with all showing similar results. Extended Data Fig. 3j was repeated twice in both sexes, with both showing similar results. Extended Data Fig. 6a was repeated in four separate cohorts in both sexes, with all showing similar results. Extended Data Figs. 8a,d (right) and 10e were repeated twice in males only, with both cohorts showing similar results. Extended Data Fig. 8b was repeated three times, with all showing similar results.

Image visualization

Brain slice schematics in Figs. 2g,i,j, 3a, 4a and 5a,b and Extended Data Figs. 4d,f,g, 7c, 8a,b,d and 10d were adapted from the Allen Brain Atlas Reference using Adobe Illustrator.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw data for animal behaviours, ISH and IHC are available as Source data files. The Allen Brain Atlas ISH database was used to search for potential molecular markers in LS.

Code availability

All MATLAB code for Ca^{2+} imaging analysis and Python code for iDISCO+ analysis can be obtained from github (https://github.com/nyclong/2021-07-11642-Nature.git).

References

1. Feder, A., Nestler, E. J. & Charney, D. S. Psychobiology and molecular genetics of resilience. *Nat. Rev. Neurosci.* **10**, 446–457 (2009).

2. Rappaport, B. I. et al. Peer victimization and dysfunctional reward processing: ERP and behavioral responses to social and monetary rewards. *Front. Behav. Neurosci.* **13**, 120 (2019).

3. Ethridge, P., Sandre, A., Dirks, M. A. & Weinberg, A. Past-year relational victimization is associated with a blunted neural response to rewards in emerging adults. *Soc. Cogn. Affect. Neurosci.* **13**, 1259–1267 (2018).

4. Berton, O. et al. Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311**, 864–868 (2006).

5. Krishnan, V. et al. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131**, 391–404 (2007).

6. Barkus, E. & Badcock, J. C. A transdiagnostic perspective on social anhedonia. *Front. Psychiatry* **10**, 216 (2019).

7. Leary, M. R. Emotional responses to interpersonal rejection. *Dialogues Clin. Neurosci.* **17**, 435–441 (2015).

8. Yizhar, O. & Levy, D. R. The social dilemma: prefrontal control of mammalian sociability. *Curr. Opin. Neurobiol.* **68**, 67–75 (2021).

9. Pegg, S. et al. Blunted social reward responsiveness moderates the effect of lifetime social stress exposure on depressive symptoms. *Front. Behav. Neurosci.* **13**, 178 (2019).

10. Golden, S. A., Covington, H. E. III, Berton, O. & Russo, S. J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* **6**, 1183–1191 (2011).

11. Takahashi, A. et al. Establishment of a repeated social defeat stress model in female mice. *Sci. Rep.* **7**, 12838 (2017).

12. Renier, N. et al. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **165**, 1789–1802 (2016).

13. Borie, A. M. et al. Correction of vasopressin deficit in the lateral septum ameliorates social deficits of mouse autism model. *J. Clin. Invest.* **131**, e144450 (2021).

14. Wong, L. C. et al. Effective modulation of male aggression through lateral septum to medial hypothalamus projection. *Curr. Biol.* **26**, 593–604 (2016).

15. Leroy, F. et al. A circuit from hippocampal CA2 to lateral septum disinhibits social aggression. *Nature* **564**, 213–218 (2018).

16. Menon, R., Suss, T., Oliveira, V. E. M., Neumann, I. D. & Bludau, A. Neurobiology of the lateral septum: regulation of social behavior. *Trends Neurosci.* **45**, 27–40 (2022).

17. Vega-Quiroga, I., Yarur, H. E. & Gysling, K. Lateral septum stimulation disinhibits dopaminergic neurons in the antero-ventral region of the ventral tegmental area: role of GABA-A alpha 1 receptors. *Neuropharmacology* **128**, 76–85 (2018).

18. Besnard, A. & Leroy, F. Top-down regulation of motivated behaviors via lateral septum sub-circuits. *Mol. Psychiatry* https://doi.org/10.1038/s41380-022-01599-3 (2022).

19. Wirtshafter, H. S. & Wilson, M. A. Lateral septum as a nexus for mood, motivation, and movement. *Neurosci. Biobehav. Rev.* **126**, 544–559 (2021).

20. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

21. Besnard, A. et al. Dorsolateral septum somatostatin interneurons gate mobility to calibrate context-specific behavioral fear responses. *Nat. Neurosci.* **22**, 436–446 (2019).

22. Azevedo, E. P. et al. A limbic circuit selectively links active escape to food suppression. *eLife* **9**, e58894 (2020).

23. Sweeney, P. & Yang, Y. An inhibitory septum to lateral hypothalamus circuit that suppresses feeding. *J. Neurosci.* **36**, 11185–11195 (2016).

24. Anthony, T. E. et al. Control of stress-induced persistent anxiety by an extra-amygdala septohypothalamic circuit. *Cell* **156**, 522–536 (2014).

25. Menon, R. et al. Oxytocin signaling in the lateral septum prevents social fear during actation. *Curr. Biol.* **28**, 1066–1078 (2018).

26. Shin, S. et al. Drd3 signaling in the lateral septum mediates early life stress-induced social dysfunction. *Neuron* **97**, 195–208 (2018).

27. Labonte, B. et al. Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).

28. Jing, W. et al. A circuit of COCH neurons encodes social-stress-induced anxiety via MTF1 activation of Cacna1h. *Cell Rep.* **37**, 110177 (2021).

29. Besnard, A., Miller, S. M. & Sahay, A. Distinct dorsal and ventral hippocampal CA3 outputs govern contextual fear discrimination. *Cell Rep.* **30**, 2360–2373 (2020).

30. Wang, D. et al. Regulation of depression-related behaviors by GABAergic neurons in the lateral septum through periaqueductal gray

neuronal projections. *J. Psychiatr. Res.* **137**, 202–214 (2021).

31. Swanson, L. W. & Cowan, W. M. The connections of the septal region in the rat. *J. Comp. Neurol.* **186**, 621–655 (1979).

32. Risold, P. Y. & Swanson, L. W. Connections of the rat lateral septal complex. *Brain Res. Rev.* **24**, 115–195 (1997).

33. Lo, L. & Anderson, D. J. A Cre-dependent, anterograde transsynaptic viral tracer for mapping output pathways of genetically marked neurons. *Neuron* **72**, 938–950 (2011).

34. Chen, P. & Hong, W. Neural circuit mechanisms of social behavior. *Neuron* **98**, 16–30 (2018).

35. Russo, S. J. & Nestler, E. J. The brain reward circuitry in mood disorders. *Nat. Rev. Neurosci.* **14**, 609–625 (2013).

36. Dolen, G., Darvishzadeh, A., Huang, K. W. & Malenka, R. C. Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* **501**, 179–184 (2013).

37. Hung, L. W. et al. Gating of social reward by oxytocin in the ventral tegmental area. *Science* **357**, 1406–1411 (2017).

38. Russo, S. J., Murrough, J. W., Han, M. H., Charney, D. S. & Nestler, E. J. Neurobiology of resilience. *Nat. Neurosci.* **15**, 1475–1484 (2012).

39. Falkner, A. L. et al. Hierarchical representations of aggression in a hypothalamic-midbrain circuit. *Neuron* **106**, 637–648 (2020).

40. Lefler, Y., Campagner, D. & Branco, T. The role of the periaqueductal gray in escape behavior. *Curr. Opin. Neurobiol.* **60**, 115–121 (2020).

41. Tovote, P. et al. Midbrain circuits for defensive behaviour. *Nature* **534**, 206–212 (2016).

42. Cezario, A. F., Ribeiro-Barbosa, E. R., Baldo, M. V. & Canteras, N. S. Hypothalamic sites responding to predator threats-the role of the dorsal

premammillary nucleus in unconditioned and conditioned antipredatory defensive behavior. *Eur. J. Neurosci.* **28**, 1003–1015 (2008).

43. Stolzenberg, D. S. & Mayer, H. S. Experience-dependent mechanisms in the regulation of parental care. *Front. Neuroendocrinol.* **54**, 100745 (2019).

44. Mahn, M. et al. Efficient optogenetic silencing of neurotransmitter release with a mosquito rhodopsin. *Neuron* **109**, 1621–1635 (2021).

45. Copits, B. A. et al. A photoswitchable GPCR-based opsin for presynaptic inhibition. *Neuron* **109**, 1791–1809 (2021).

46. Krach, S., Paulus, F. M., Bodden, M. & Kircher, T. The rewarding nature of social interactions. *Front. Behav. Neurosci.* **4**, 22 (2010).

47. Trezza, V., Campolongo, P. & Vanderschuren, L. J. Evaluating the rewarding nature of social interactions in laboratory animals. *Dev. Cogn. Neurosci.* **1**, 444–458 (2011).

48. Lee, C. R., Chen, A. & Tye, K. M. The neural circuitry of social homeostasis: consequences of acute versus chronic social isolation. *Cell* **184**, 2794–2795 (2021).

49. Kupferberg, A., Bicks, L. & Hasler, G. Social functioning in major depressive disorder. *Neurosci. Biobehav. Rev.* **69**, 313–332 (2016).

50. McLaughlin, K. A., Colich, N. L., Rodman, A. M. & Weissman, D. G. Mechanisms linking childhood trauma exposure and psychopathology: a transdiagnostic model of risk and resilience. *BMC Med.* **18**, 96 (2020).

51. Hodes, G. E. et al. Sex differences in nucleus accumbens transcriptome profiles associated with susceptibility versus resilience to subchronic variable stress. *J. Neurosci.* **35**, 16362–16376 (2015).

52. Gunaydin, L. A. et al. Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535–1551 (2014).

53. Zernig, G., Kummer, K. K. & Prast, J. M. Dyadic social interaction as an alternative reward to cocaine. *Front. Psychiatry* **4**, 100 (2013).

54. Flanigan, M. E. et al. Orexin signaling in GABAergic lateral habenula neurons modulates aggressive behavior in male mice. *Nat. Neurosci.* **23**, 638–650 (2020).

55. Denninger, J. K., Smith, B. M. & Kirby, E. D. Novel object recognition and object location behavioral testing in mice on a budget. *J. Vis. Exp.* **20**, 10.3791/58593 (2018).

56. Angoa-Perez, M., Kane, M. J., Briggs, D. I., Francescutti, D. M. & Kuhn, D. M. Marble burying and nestlet shredding as tests of repetitive, compulsive-like behaviors in mice. *J. Vis. Exp.* **24**, 50978 (2013).

57. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

58. Lee, H. et al. Scalable control of mounting and attack by Esr1$^+$ neurons in the ventromedial hypothalamus. *Nature* **509**, 627–632 (2014).

59. Martianova, E., Aronson, S. & Proulx, C. D. Multi-fiber photometry to record neural activity in freely-moving animals. *J. Vis. Exp.* https://doi.org/10.3791/60278 (2019).

Acknowledgements

We thank A. Smith for help with the iDISCO+ protocol and ClearMap analysis, M. Flanigan for advice on validation of the sCPP test, N. Tzavaras, K. Cialowicz and G. Doherty from Mount Sinai Microscopy CoRE for help with lightsheet imaging and N. Rome and S. Gaydos for help with brain slicing. This research was supported by US National Institutes of Health grant nos. R01MH114882, R01MH127820, R01MH104559, R01MH120514 and R01MH120637 (S.J.R.), Brain & Behavior Research Foundation (no. 30233 to L.L., no. 29104 to F.C. and no. 29699 to C.M.), Canadian Institutes of Health Research (no. 201811MFE-414896-231226 to K.L.C.) and NIH Virus Center (grant no. P40 OD010996).

Author information

Authors and Affiliations

1. Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

   Long Li, Romain Durand-de Cuttoli, Antonio V. Aubry, C. Joseph Burnett, Flurin Cathomas, Lyonna F. Parise, Kenny L. Chan, Yusuke Shimo & Scott J. Russo

2. Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

   Long Li, Romain Durand-de Cuttoli, Antonio V. Aubry, C. Joseph Burnett, Flurin Cathomas, Lyonna F. Parise, Kenny L. Chan, Carole Morel, Chongzhen Yuan, Yusuke Shimo, Hsiao-yun Lin, Jun Wang & Scott J. Russo

3. Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

   Carole Morel

4. Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

   Chongzhen Yuan, Hsiao-yun Lin & Jun Wang

5. James J Peters VA Medical Center, Research & Development, New York, NY, USA

   Chongzhen Yuan, Hsiao-yun Lin & Jun Wang

Contributions

L.L. and S.J.R. conceived the project. Surgeries were performed by L.L., L.F.P., C.M. and R.D.C. Tissue preparation was performed by L.L., C.J.B.,

F.C., L.F.P., K.L.C., Y.S. and H.-Y.L. IHC/FISH, microscopy imaging/analysis, fibre photometry data collection and analysis were performed by L.L. Behavioural experiments were performed by L.L. and C.Y. Electrophysiology experiments were performed by R.D.C. iDISCO+ procedures were performed by L.L. iDISCO+ data were analysed by A.V.A. J.W. provided intellectual input and edited the manuscript. Results were analysed and interpreted by L.L., R.D.C., A.V.A. and S.J.R. The manuscript was written by L.L., R.D.C., A.V.A. and S.J.R. and edited by all authors.

Corresponding author

Correspondence to Scott J. Russo.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Dayu Lin, Moriel Zelikowsky and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Detailed ethological analysis of social behaviour alterations in CSDS animals.

a—**f**, SUS mice show higher corner ratio (One-Way ANOVA, female, F (2, 31) = 29.62, $P < 0.0001$, $n = 10$ (CTRL), 12 (RES), 12 (SUS) (**a**), males, F (2, 46) = 17.58, $P < 0.0001$, $n = 10$ (CTRL), 13 (RES), 26 (SUS) (**d**)), while showing no locomotor activity deficits (One-Way ANOVA, female, F (2, 31) = 0.8416, $P = 0.4406$ (**b**), males, F (2, 46) = 0.8416, $P = 0.4376$, (**e**)) during SI test and showing longer latency to first social bout (One-Way ANOVA, female, F (2, 31) = 8.399, $P = 0.0012$, (**c**), males, F (2, 46) = 16.63, $P < 0.0001$, (**f**)) during RI test. **g**—**l**, Correlation between SI ratio and social investigation time (female, $R^2 = 0.1728$, $P = 0.0145$ (**g**), male, $R^2 = 0.1958$, $P = 0.0015$ (**j**)), social avoidance (female, $R^2 = 0.3399$, $P = 0.0003$ (**h**), male, $R^2 = 0.1847$, $P = 0.0021$ (**k**)) and latency to first social bout (female, $R^2 = 0.1464$, $P = 0.0255$ (**i**), male, $R^2 = 0.1366$, $P = 0.0090$ (**l**)). **m**—**r**, Correlation between CPP score and social investigation time (female, $R^2 = 0.2818$, $P = 0.0012$ (**m**), male, $R^2 = 0.1533$, $P = 0.0054$ (**p**)), social avoidance (female, $R^2 = 0.1995$, $P = 0.0081$ (**n**), male, $R^2 = 0.0911$, $P = 0.0351$ (**q**)) and latency to first social bout (female, $R^2 = 0.3065$, $P = 0.0007$ (**o**), male, $R^2 = 0.0318$, $P = 0.2200$ (**r**)). **s**, Subtracted CPP score of different stages of female estrus cycle show stage of the cycle has no effect on social CPP formation (One-Way ANOVA, F (3, 20) = 0.5148, $P = 0.6768$, $n = 4$ (Proestrus), 6 (Estrus), 4 (Metestrus), 10 (Diestrus)). **t**, Subtracted CPP score of different social targets for female social CPP. **u**, Distribution of different behaviour parameters of CTRL, RES, SUS animals. ns, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. All data are expressed as mean ± s.e.m.

Source data

Extended Data Fig. 2 Memory tests in animals exposed to CSDS, effects of reversed test order on social behaviours, and behavioural parameters grouped by CPP score.

a, Schematic diagram of novel object recognition (NOR) test and novel location test (NLT). **b**, Novel object investigation time (One-Way ANOVA,

F (2, 29) = 3.041, *P* = 0.0633, *n* = 10 (CTRL), 9 (RES), 13 (SUS)) and **c**, novel location investigation time (One-Way ANOVA F (2, 29) = 0.5601, *P* = 0.5772). **d**, Schematic diagram of reversed behaviour test order. **e**, Subtracted CPP score (One-Way ANOVA, F (2, 22) = 3.984, *P* = 0.0334), social investigation time (F (2, 22) = 8.267, *P* = 0.0021), social avoidance (F (2, 22) = 9.919, *P* = 0.0008) and SI ratio (F (2, 22) = 18.32, *P* < 0.0001, n = 8 (CTRL), 7 (RES), 9 (SUS). **f**, **g**, Social behavioural parameters after grouping by CPP score, (**f**) females, SI ratio (two-tailed t-test, t = 1.660, df = 32, *P* = 0.1067), social investigation time (t = 3.788, df = 32, *P* = 0.0006), social avoidance (t = 3.001, df = 32, *P* = 0.0052), latency (t=3.204, df = 32, *P* = 0.0031), n = 11 (CPP-), 23 (CPP+). (**g**) males, SI ratio (two-tailed t-test, t = 2.298, df = 47, *P* = 0.0261), social investigation time (t=2.868, df = 47, *P* = 0.0062), social avoidance (t = 2.545, df = 47, *P* = 0.0143), latency (t = 1.438, df = 47, *P* = 0.1570), n = 21 (CPP-), 28 (CPP+). ns, not significant, * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001. All data are expressed as mean ± s.e.m.

[Source data]

[Extended Data Fig. 3 Cleared whole brain activity data.]

a—**d**, Female iDISCO+ groups' SI ratio (One-Way ANOVA, F (2, 27) = 22.25, *P* < 0.0001, *n* = 8 (CTRL), 11 (RES), 11 (SUS)) (**a**), social investigation time (One-Way ANOVA, F (2, 27) = 20.24, *P* < 0.0001) (**b**), social avoidance (One-Way ANOVA, F (2, 27) = 7.747, *P* = 0.0022) (**c**) and latency to first social bout (One-Way ANOVA, F (2, 27) = 6.075, *P* = 0.0066) (**d**). **e**—**h**, Male iDISCO+ groups' SI ratio (One-Way ANOVA, F (2, 25) = 13.85, *P* < 0.0001, *n* = 9 (CTRL), 11 (RES), 8 (SUS)) (**e**), social investigation time (One-Way ANOVA, F (2, 25) = 14.07, *P* < 0.0001) (**f**), social avoidance (One-Way ANOVA, F (2, 25) = 38.76, *P* < 0.0001) (**g**) and latency to first social bout (One-Way ANOVA, F (2, 25) = 7.370, *P* = 0.0030) (**h**). **i**, ClearMap cFos detection verification shows annotation (green dots) matching cFos signal (red nuclei). **j**, Brain slice cFos staining verification in males showing most of the $cFos^+$ cells are in the lateral part of the lateral septum. **k**, Statistics of LS cFos during RI test following CSDS in males (Unpaired two-tailed t-test, CTRL, t4 = 1.434, *P* = 0.2248, *n* = 3 per group, RES, t4 = 1.957, *P* = 0.1219, *n* = 3 per group, SUS, t9 = 4.429, *P* =

0.0016, n = 5 (object), 6 (intruder)). ns, not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. All data are expressed as mean ± s.e.m. Scale bar, 100 µm.

Source data

Extended Data Fig. 4 Characterization of co-expression of neurotensin and other genes in lateral septum.

a, Ratio of $Nt^{+}cFos^{+}/Nt^{+}$ (492 of 520 neurons) and $Nt^{+}cFos^{+}/cFos^{+}$ (492 of 496 neurons) in LS of SUS mice. **b**,**c**, Multiplex ISH shows *Nt* and *Gad2* colocalization in anterior, middle and posterior LS (**b**) shows most of the neurotensin neurons are GABAergic neurons (**c**). (n = 3 slices per mouse, n = 3 mice per group, scale bar, 100 µm). **d**,**e**, Multiplex ISH shows *Nt* and *Crhr2* colocalization in LS (**d**) shows *Nt* neurons and *Crhr2* neurons are largely overlapping in anterior and middle parts of the LS, but show very low colocalization in the posterior part of the LS (**e**). (n = 3 slices per mouse, n = 3 mice per group, scale bar, 100 µm). **f**, **g**, Multiplex ISH shows *Nt* and *Drd3* do not overlap in the LS (n = 3 slices per mouse, n = 3 mice per group, scale bar, 50 µm). **h**, **i**, Multiplex ISH shows *Nt* and *Oxtr* showing very low overlap in the LS (n = 3 slices per mouse, n = 3 mice per group, scale bar, 50 µm). **j**, Confocal images of *Sst* and cFos expression in CTRL, RES and SUS mice after RI test. **k**, Comparison of Sst^{+} $cFos^{+}$ neurons between the three groups (One-Way ANOVA, $F(2, 9) = 4.780$, $P = 0.0385$, n = 4 per group, Scale bar, 100 µm). **l**, Comparison of Nt^{-} $cFos^{+}$ neurons between female CTRL, RES and SUS mice (One-Way ANOVA, $F(2, 6) = 0.2755$, $P = 0.7683$, n = 3 per group) and males ($F(2, 10) = 4.980$, $P = 0.0316$, n = 3–6). ns, not significant, * $P < 0.05$. All data are expressed as mean ± s.e.m.

Source data

Extended Data Fig. 5 Slice electrophysiology following CSDS, measurement of Ca^{2+} activity in NT^{LS} neurons responses during food reward and social conditioning.

a, Characterization of *ex vivo* electrophysiology parameters measured in NT^{LS} neurons following CSDS in SUS and RES mice. From left to right: Bar graph of average resting membrane potential per neuron (mean SUS= −67.49 mV +/− 0.975 mV, *n* = 55 with 4–8 neurons/mouse; mean RES = −61.62 mV +/− 1.73 mV, *n* = 19 with 4–7 neurons/mouse; two-tailed Welch's t test, *P* = 0.0059). No difference in action potential threshold (two-tailed Welch's test, *P* = 0.3037), amplitude (two-tailed Welch's test, *P* = 0.6661), half-width duration (two-tailed Welch's test, *P* = 0.3757) or fast after hyperpolarization (fAHP) amplitude (two-tailed Welch's test, *P* = 0.7154). **b**, Social interaction ratio of fiber photometry cohort after CSDS in females (One-Way ANOVA, F (2, 11) = 5.629, *P* = 0.0207, *n* = 4 (CTRL), 5 (RES), 5 (SUS)) and males (One-Way ANOVA, F (2, 14) = 12.93, P = 0.0007, *n* = 7 (CTRL), 5 (RES), 5 (SUS (5)). **c**, Ca^{2+} activity in NT^{LS} neurons during consumption of a peanut butter cup (left). Dashed line in **c** marks the beginning of biting. Comparison of Ca^{2+} ΔF/F change before and after biting (right, two-tailed paired t test, t = 1.577, df = 4, *P* = 0.1900, *n* = 5). **d**, Ca^{2+} traces in paired (pink) and unpaired (blue) conditioning chambers. The inset shows normalized area under curve (AUC) between paired and unpaired conditioning sessions (sexes combined, paired two-tailed t-test, CTRL, t = 0.1977, df = 5, *P* = 0.8511, RES, t = 1.049, df = 5, *P* = 0.3423, SUS, t = 5.453, df = 5, *P* = 0.0028), *n* = 6 per group). ns, not significant, * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001. All data are expressed as mean ± s.e.m.

Source data

Extended Data Fig. 6 NT^{LS} neuron manipulation in stress-naïve mice does not change social interaction and CVS has no effect on sCPP.

a, Expression of AAV-DIO-DREADDs in NT^{LS} neurons. **b**, NT^{LS} activation in stress-naïve mice does not change social behaviour in females (unpaired two-tailed t-test, t14 = 1.044, *P* = 0.3141, *n* = 8 per group) or males (unpaired two-tailed t-test, t14 = 1.434, *P* =0.3975, *n* = 8 per group). NT^{LS} activation in stress-naïve mice does not change locomotor activity in females (unpaired two-tailed t-test, t14 = 0.3473, *P* = 0.7335, *n* = 8 per group) or males (unpaired two-tailed t-test, t14 = 1.425, *P* = 0.1762, *n* = 8 per group). **c**, NT^{LS}

chemogenetic activation in stress-naïve mice does not change social preference (Two-way repeated measures ANOVA, Vehicle, F = (1, 16) = 7.198, *P* = 0.0163, *n* = 9, CNO, F (1, 18) = 6.644, *P* = 0.0190, *n* = 10). **d**, Subtracted CPP score comparison between vehicle and CNO group (unpaired two-tailed t-test, t = 0.03518, df = 17, *P* = 0.9723, *n* = 10 per group). **e**, Schematic diagram of CVS and CPP test. **f**, CVS effect on sCPP in either sex (Two-way repeated measures ANOVA, female, CTRL, F (1, 22) = 4.824, *P* = 0.0389, *n* = 12, CVS, F (1, 22) = 5.172, *P* = 0.0331, *n* = 12; male, CTRL, F (1, 22) = 5.042, *P* = 0.0351, *n* = 12; CVS, F (1, 22) = 3.900, *P* = 0.0610, *n* = 12). **g**, Schematic diagram of virus injection and optogenetic manipulation of NT^{LS} neurons during RTPP test. **h**, NT^{LS} neuron activation does not alter real time place preference in either stress-naïve females (Paired two-tailed t-test, ChR2, t11 = 0.06179, *P* = 0.9518, *n* = 12, EYFP, t15 = 0.01923, *P* = 0.9849, *n* = 16) or males (Paired two-tailed t-test, ChR2, t8 = 0.3855, *P* = 0.7099, *n* = 9; EYFP, t8 = 0.6333, *P* = 0.5442, *n* = 9). ns, not significant. * *P* < 0.05. All data are expressed as mean ± s.e.m. Scale bar, 100 μm. BioRender was used to generate schematic figures in **c**,**e**,**g**.

Source data

Extended Data Fig. 7 NT^{LS} neurons modulate chronic restraint stress-induced object avoidance and anxiety-like behaviours.

a, Schematic diagram of chronic restraint stress and RI test. **b**, Comparison of latency to the first investigation (unpaired two-tailed t-test, t = 3.142, df = 18, *P* = 0.0056) and number of investigation bouts (t = 5.259, df = 18, *P* < 0.0001) with a novel restraint tube, or withdrawal from the tube (t = 5.229, df = 18, *P* < 0.0001), *n* = 10, between control (CTRL) and chronic restraint stressed (CRS) animals. **c**, Schematic diagram of CRS with DREADD manipulation during RI test. **d**, Latency (paired two-tailed t-test, t = 2.065, df = 8, *P* = 0.0728), and number of investigation bouts (t = 2.619, df = 8, *P* = 0.0307) with a novel restraint tube, or withdrawal from the tube (t = 1.988, df = 8, *P* = 0.0820), *n* = 9. **e**, Schematic diagram of chronic restraint stress with juvenile bedding/odour (CRSO) and RI test. **f**, Latency (unpaired two-tailed t-test, t = 0.5452, df = 18, *P* = 0.5923) and number of investigation bouts (t = 0.3971, df = 18, *P* = 0.6960) with a novel restrainer tube, or withdrawal from the tube (t = 0.000, df = 18, *P* > 0.9999), *n* = 10. **g**, NT^{LS}

activation in stress-naïve male mice leads to higher anxiety-like behaviour in the elevated plus maze (unpaired two-tailed t-test, $t14 = 2.824$, $P = 0.0135$, $n = 8$ per group). **h**, NT^{LS} activation or inhibition in stress-naïve male mice modulate marble burying behaviour (two-tailed paired t-test, hM3Dq, $t7 = 4.631$, $P = 0.0024$, $n = 8$; hM4Di, $t7 = 4.020$, $P = 0.0051$, $n = 8$). **i**, NT^{LS} activation or inhibition in stress-naïve male mice leads to higher or lower anxiety levels, respectively, in the open field test (unpaired two-tailed t-test, hM3Dq, $t14 = 2.189$, $P = 0.0461$, $n = 8$ per group; hM4Di, $t14 = 1.424$, $P = 0.1762$, $n = 8$ per group). **j**, with no locomotor changes (unpaired two-tailed t-test, hM3Dq, $t14 = 1.641$, $P = 0.1230$; hM4Di, $t14 = 1.566$, $P = 0.1398$). ns, not significant. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$. All data are expressed as mean ± s.e.m. BioRender was used to generate schematic figures in **a**,**c**,**e**,**i**.

Source data

Extended Data Fig. 8 NT^{LS} monosynaptic downstream region verification.

a, Schematic diagram of H129ΔTK-TT virus injection. 48 h after injection, there were no $tdTomato^{+}$ neurons in downstream regions of NT^{LS} neurons. **b**, Retrograde AAV-DIO-EGFP/tdTomato tracing verification shows NT^{LS}→NAc, NT^{LS}→AHN, NT^{LS}→PAG monosynaptic connections. **c**, Colocalization analysis for overlapping NT^{LS} projection neurons to the NAc/AHN/PAG. **d**, Cholera toxin subunit b (CTB) tracing for NT^{LS}→NAc, NT^{LS}→AHN, and NT^{LS}→PAG projections. Representative image of LS (middle panel) with yellow arrowheads indicating neurons projecting to both NAc and AHN. White arrowheads indicate neurons projecting to both PAG and AHN. **e**, Number of projections showing colocalization of CTB tracers from three downstream regions (3 slices per brain region per mouse, $n = 3$ mice). All data are expressed as mean ± s.e.m. Scale bar, 100 µm.

Extended Data Fig. 9 Role of NT^{LS} inputs to NAc and AHN in regulating social behaviours.

a, Schematic of NT^{LS} targeting with NpHR and manipulation during social behaviour tests. **b**–**e**, NpHR axon terminal inhibition in the NAc (**b**, **c**) and AHN (**d**, **e**) rescued social investigation time and partially rescued social avoidance in both females (**b**, social investigation, F (1, 14) = 3.484, *P* = 0.0831, *n* = 8 per group; social avoidance, F (1, 14) = 1.180, *P* = 0.2956, *n* = 8 per group, **d**, social investigation, F (1, 14) = 4.982, *P* = 0.0425, *n* = 8 per group; social avoidance, F (1, 14) = 2.266, *P* = 0.1545, *n* = 8 per group) and males (**c**, social investigation, F (1, 14) = 0.2046, *P* = 0.6580, *n* = 8 per group; social avoidance, F (1, 14) = 1.214, *P* = 0.2891, *n* = 8 per group, **e**, social investigation, F (1, 14) = 4.597, *P* = 0.0501, *n* = 8 per group; social avoidance, F (1, 14) = 5.359, *P* = 0.0363, *n* = 8 per group). **f**-**m**, Optogenetic manipulation of NT^{LS}→NAc and NT^{LS}→AHN circuits during social CPP conditioning in both sexes. **f**, Activation of NT^{LS}→NAc with ChR2 blocked social reward in RES females (EYFP, F (1, 12) = 2.362, *P* = 0.1502, *n* = 7, ChR2, F (1, 14) = 0.5543, *P* = 0.4689, *n* = 8) **g**, Inhibition of NT^{LS}→NAc with NpHR rescued social reward deficits in SUS females (EYFP, F (1, 14) = 0.5105, *P* = 0.4867, NpHR, F (1, 14) = 4.183, *P* = 0.0601, *n* = 8 per group). **h**, Activation of NT^{LS}→AHN with ChR2 blocked social reward in RES females (EYFP, F (1, 12) = 4.289, *P* = 0.0606, *n* = 7, ChR2, F (1, 14) = 0.0001296, *P* = 0.9911, *n* = 8). **i**, Inhibition of NT^{LS}→AHN with NpHR rescued social reward deficits in SUS females (EYFP, F (1, 14) = 0.1068, *P* = 0.7487, NpHR, F (1, 14) = 4.575, *P* = 0.0505, *n* = 8 per group). **j**, Activation of NT^{LS}→NAc with ChR2 blocked social reward in RES males (EYFP, F (1, 12) = 5.703, *P* = 0.0343, *n* = 7, ChR2, F (1, 14) = 0.2484, *P* = 0.6259, *n* = 8). **k**, Inhibition of NT^{LS}→NAc with NpHR rescued social reward deficits in SUS males (EYFP, F (1, 14) = 4.053, *P* = 0.0637, NpHR, F (1, 14) = 4.140, *P* = 0.0613, *n* = 8 per group). **l**, Activation of NT^{LS}→AHN with ChR2 blocked social reward in RES males (EYFP, F (1, 12) = 6.329, *P* = 0.0271, *n* = 7, ChR2, F (1, 14) = 0.1567, *P* = 0.6981, *n* = 8). **m**, Inhibition of NT^{LS}→AHN with NpHR rescued social reward deficits in SUS males (EYFP, F (1, 12) = 0.6881, *P* = 0.4230, NpHR, F (1, 14) = 10.02, *P* = 0.0069, *n* = 8 per group). Two-way repeated measures ANOVA were performed for any comparison in this figure: ns, not significant. * $P < 0.05$, ** $P < 0.01$. All data are expressed as mean ± s.e.m. BioRender was used to generate schematic figures in **a**.

Source data

Extended Data Fig. 10 Optogenetic validation of NT^{LS} monosynaptic projections and effects of optogenetic stimulation of non-NT^{LS} neurons on social behaviour.

a,**b**, ChR2-assisted circuit mapping of NT^{LS}→NAc and NT^{LS}→AHN pathways showing monosynaptic (with TTX), inhibitory (Cs-based internal, clamped at 0 mV) and GABAa-dependent (SR-95531, Gabazine) connections in 5/8 NAc/NDB neurons and 4/10 AHN neurons. **c**, Light-evoked inward current in a ChR2-$EYFP^+$ Cre^+ neuron in voltage-clamp mode (clamped at −70 mV) upon 1 s illumination with 470 nm blue light (left) and 15 Hz pulse-induced spikes (right). **d**, Schematic diagram of non-NT^{LS} neuron infection and manipulation during social behaviour tests. **e**,**f**, Multiplex ISH for *Nt* and CreOff-ChR2-EYFP. **c**, Overlap of Nt^+ $EYFP^+$ neurons among $EYFP^+$ neurons (2-3 slices per mouse, $n = 5$ mice). **g**, Light-evoked inward current in ChR2-$EYFP^+$ non-Cre neurons in voltage-clamp mode (clamped at −70 mV) upon 1 s illumination with blue light (470 nm). **h**, Effect of ChR2 stimulation of non-*Nt* neurons on social investigation and social avoidance during RI test (Mixed-effects analysis two-way ANOVA, left, $F(1, 13) = 0.2137$, $P = 0.6516$, right, $F(1, 13) = 0.5672$, $P = 0.4648$, n = ChR2 (10), EYFP (5)). ns, not significant. All data are expressed as mean ± s.e.m. Scale bar, 50 µm.

Source data

Extended Data Table 1 Regions that show significant difference between SUS and RES females (iDISCO+ analysis)

Extended Data Table 2 Regions that show significant difference between SUS and CTRL females (iDISCO+ analysis)

Extended Data Table 3 Regions that show significant difference between RES and CTRL females (iDISCO+ analysis)

Extended Data Table 4 Regions that show significant difference between SUS and RES males (iDISCO+ analysis)

Extended Data Table 5 Regions that show significant difference between RES and CTRL males (iDISCO+ analysis)

Supplementary information

Supplementary Table 1

Regions showing significant difference between SUS and CTRL males (iDISCO+ analysis).

Reporting Summary

Peer Review File

Source data

Source Data Fig. 1.

Source Data Fig. 2.

Source Data Fig. 3.

Source Data Fig. 4.

Source Data Fig. 5.

Source Data Extended Data Fig. 1.

Source Data Extended Data Fig. 2.

Source Data Extended Data Fig. 3.

Source Data Extended Data Fig. 4.

Source Data Extended Data Fig. 5.

Source Data Extended Data Fig. 6.

Source Data Extended Data Fig. 7.

[Source Data Extended Data Fig. 9.]

[Source Data Extended Data Fig. 10.]

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Li, L., Durand-de Cuttoli, R., Aubry, A.V. *et al.* Social trauma engages lateral septum circuitry to occlude social reward. *Nature* **613**, 696–703 (2023). https://doi.org/10.1038/s41586-022-05484-5

- Received: 19 July 2021
- Accepted: 25 October 2022
- Published: 30 November 2022
- Issue Date: 26 January 2023

- DOI: https://doi.org/10.1038/s41586-022-05484-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Avoiding social interactions**
  - Katherine Whalley

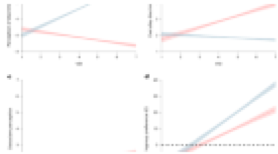
  Nature Reviews Neuroscience (2023)

Avoiding social interactions

- Katherine Whalley

Nature Reviews Neuroscience Research Highlight 21 Dec 2022

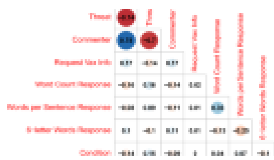
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05484-5



The association between vaccination status identification and societal polarization

03 November 2022

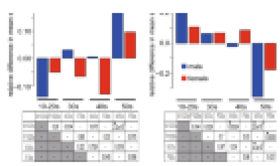
Luca Henkel, Philipp Sprengholz, … Robert Böhm



The vaccination rift effect provides evidence that source vaccination status determines the rejection of calls to get vaccinated

08 November 2022

J. Lukas Thürmer & Sean M. McCrea



A message of the majority with scientific evidence encourages young people to show their prosocial nature in COVID-19 vaccination

01 December 2021

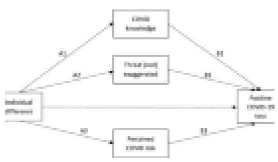
Toshiko Tanaka, Tsuyoshi Nihonsugi, … Masahiko Haruno



[Understanding hesitancy with revealed preferences across COVID-19 vaccine types](#)

02 August 2022

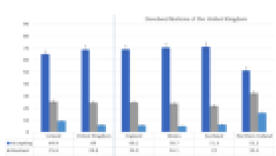
Kristóf Kutasi, Júlia Koltai, … Balázs Lengyel



[Contracting COVID-19: a longitudinal investigation of the impact of beliefs and knowledge](#)

14 October 2021

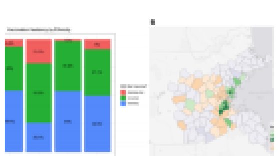
Courtney A. Moore, Benjamin C. Ruisch, … Russell H. Fazio



[Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom](#)

04 January 2021

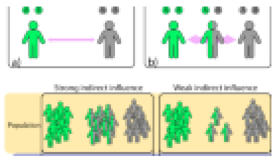
Jamie Murphy, Frédérique Vallières, … Philip Hyland



Vaccination intentions generate racial disparities in the societal persistence of COVID-19

07 October 2021

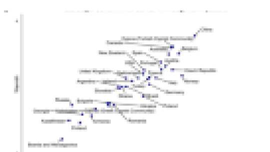
Yanchao Wang, Alina Ristea, … Daniel T. O’Brien



Mapping the global opinion space to explain anti-vaccine attraction

19 May 2022

Dino Carpentras, Adrian Lüders & Michael Quayle



Warmth and competence perceptions of key protagonists are associated with containment measures during the COVID-19 pandemic: Evidence from 35 countries

08 December 2022

Maria-Therese Friehs, Patrick F. Kotzur, … Vincent Yzerbyt

- Article
- Published: 08 December 2022

Discriminatory attitudes against unvaccinated people during the pandemic

- Alexander Bor ,
- Frederik Jørgensen &
- Michael Bang Petersen

Nature **volume 613**, pages 704–711 (2023)

- 109k Accesses

- 4023 Altmetric

- Metrics details

Abstract

During the COVID-19 pandemic, sizeable groups of unvaccinated people persist even in countries with high vaccine access[1]. As a consequence, vaccination became a controversial subject of debate and even protest[2]. Here we assess whether people express discriminatory attitudes in the form of negative affectivity, stereotypes and exclusionary attitudes in family and political settings across groups defined by COVID-19 vaccination status. We quantify discriminatory attitudes between vaccinated and unvaccinated citizens in 21 countries, covering a diverse set of cultures across the world. Across three conjoined experimental studies ($n = 15{,}233$), we demonstrate that vaccinated people express discriminatory attitudes towards unvaccinated individuals at a level as high as discriminatory attitudes that are commonly aimed at immigrant and minority populations[3,4,—5]. By contrast, there is an absence of evidence that unvaccinated individuals display discriminatory attitudes towards vaccinated people, except for the presence of negative affectivity in Germany and the USA. We find evidence in support of discriminatory attitudes against unvaccinated individuals in all

countries except for Hungary and Romania, and find that discriminatory attitudes are more strongly expressed in cultures with stronger cooperative norms. Previous research on the psychology of cooperation has shown that individuals react negatively against perceived 'free-riders'[6,7], including in the domain of vaccinations[8,9]. Consistent with this, we find that contributors to the public good of epidemic control (that is, vaccinated individuals) react with discriminatory attitudes towards perceived free-riders (that is, unvaccinated individuals). National leaders and vaccinated members of the public appealed to moral obligations to increase COVID-19 vaccine uptake[10,11], but our findings suggest that discriminatory attitudes—including support for the removal of fundamental rights—simultaneously emerged.

Main

In a historical feat of science, highly effective vaccines against SARS-CoV-2 were developed, tested, approved and mass produced in less than a year[12]. However, it soon became clear that achieving a sufficiently high uptake of these vaccines was in itself a major challenge[13]. Despite targeted vaccine mandates, vaccine passports and massive information campaigns, sizeable groups in several countries across the world continued to refuse to get vaccinated against COVID-19, even in regions in which vaccines were widely available[1]. At the same time, many countries continued to use interventions to control infection spread, resulting in feelings of pandemic fatigue, waning support for restrictions and dwindling trust in authorities[14,15,16].

Against this backdrop, public debates around COVID-19 have been heated. Some politicians have justified strict policies against unvaccinated individuals using highly moralistic rhetoric[10]. At the same time, disruptive public protests directed against vaccine mandates have taken place in several Western countries[2]. Survey research shows that divisions based on vaccination status are also emerging among the public[17,18]. Individuals who comply with the advice of health authorities morally condemn unvaccinated individuals for violating a social contract in the midst of a crisis[8,9,11]. Those who refuse vaccines report that they feel discriminated against[18] and pressured against their will[19]. Furthermore, vaccination status is consistently

aligned with other political opinions such as trust in science and the authorities and, in the case of the USA, partisanship[9,13,20].

Previous research shows that political divides can harm everyday interactions between citizens by eliciting general antipathy in the form of prejudice[21]. Here we provide a cross-cultural empirical investigation of the nature and level of prejudice across groups defined by COVID-19 vaccination status, covering in total 21 countries across all inhabited continents. We follow ref. [22] and define prejudice as “a negative evaluation of an individual that is significantly based on the individual’s group membership” (see also refs. [23,24]). Prejudice can manifest itself in affective (for example, negative emotions), cognitive (for example, negative stereotypes) and attitudinal (for example, support for exclusion and discrimination) expressions of prejudiced individuals[25]. Here we investigate all three dimensions in the context of groups defined by COVID-19 vaccination status.

Research on the psychology of vaccination decisions before the COVID-19 pandemic[8] and before the implementation of COVID-19 vaccines[9] has shown that generosity in two-player behavioural economic games is indeed affected by the vaccination status of the players. Specifically, vaccinated individuals are less generous towards unvaccinated individuals but, importantly, unvaccinated individuals are not less generous towards vaccinated individuals. These findings are interpreted on the basis of the psychology of human cooperation[8]. Research on cooperation has provided strong evidence that people monitor cooperative situations for the existence of free-riders (that is, individuals who benefit from the cooperation without paying appropriate costs)[26] and react negatively towards free-riders after detection[6,7]. Vaccinations contribute to the public good of epidemic control[27], and refusal to receive a vaccination is accordingly spontaneously perceived as an instance of free-riding, motivating contributors (that is, vaccinated individuals) to withhold generosity from unvaccinated individuals[8]. As the spontaneous withholding of resources from unvaccinated individuals may incentivize vaccination, health communicators have been advised that “making the social contract explicit may help to increase vaccine uptake rates without relying on mandates”[8]. On most normative grounds, it is unproblematic if people—as shown in previous

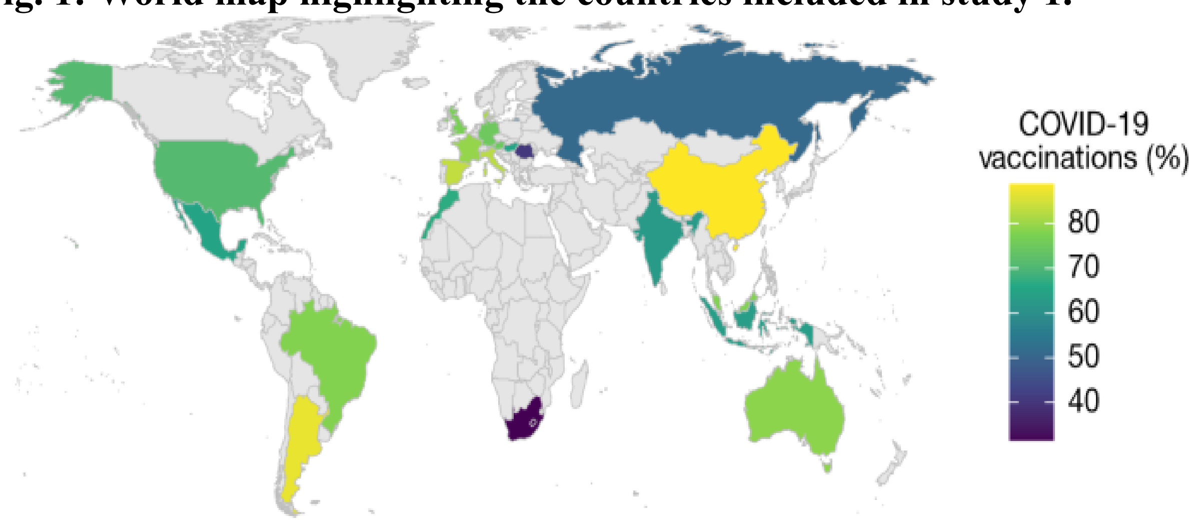
studies—are generous only towards cooperators and withhold personal resources from strangers who are known to free-ride[28].

Yet, in highly polarized contexts such as vaccinations during the COVID-19 pandemic, it is possible that these psychological processes shift in multiple important ways beyond the findings of previous research on vaccination status and generosity. First, research on the psychology of cooperation suggests that two distinct psychological motivations are activated in the context of public goods provisions: motivations to generously offer rewards to contributors and motivations to punitively impose costs on free-riders[6]. Although previous research focused on the former, it is plausible that the polarized and moralized sentiments surrounding COVID-19 vaccination activate the latter, punitive, motivations too. Thus, vaccinated people may not only suspend their generosity towards unvaccinated individuals, but may also express support for the imposition of costs on unvaccinated individuals by, for example, supporting their exclusion from social relationships or democratic rights and freedoms. Second, in this context, unvaccinated individuals may react with prejudice towards vaccinated individuals as well, grounded, for example, in perceived pressure and discrimination[18,19]. Indeed, a study examining generosity in two-player behavioural economic games after the implementation of COVID-19 vaccines found that unvaccinated individuals were also less generous towards vaccinated individuals, although ingroup favouritism was smaller than among vaccinated individuals[18]. Third, the complexity of the debates surrounding COVID-19 vaccinations may fuel negative stereotypes beyond the dimensions most relevant to cooperative dilemmas. For example, research on impression formation documents that warmth is one major dimension of impression formation, which is directly related to cooperativeness[29]. Consistent with this, research before the COVID-19 pandemic found that vaccinated individuals perceive unvaccinated individuals as less warm[8]. However, research on impression formation also documents that impressions of competence constitute another and independent evaluative dimension[29]. In the context of COVID-19 vaccines, this other dimension may also be activated as, for example, vaccinated individuals may perceive unvaccinated individuals as being unintelligent and incompetent for believing false information regarding vaccinations[30]. Discriminatory attitudes in the context

of COVID-19 vaccines may therefore come to have a broader cognitive basis.

To empirically examine these possibilities, we leverage large-scale cross-national data. Specifically, we conducted three experimental studies in 21 countries (study 1; *n* = 64,440 observations from 10,740 respondents); six countries (study 2; *n* = 18,270 observations from 3,045 respondents); and in the USA (study 3; *n* = 14,480 observations from 1,448 respondents), studying the affective, cognitive and attitudinal dimensions of prejudice across groups defined by COVID-19 vaccination status. The dataset measures discriminatory attitudes across a diverse set of cultures from all inhabited continents of the world (Fig. 1). As previous research on the lack of generosity towards unvaccinated individuals has been limited to Western democratic contexts[8,9,18], this cross-cultural dataset sheds light on both the ubiquity of discriminatory attitudes against perceived free-riders as well as on the cross-cultural predictors of variation in the strength of such attitudes. If discriminatory attitudes against people who are not vaccinated against COVID-19 reflect the activation of anti-free-rider sentiments, such attitudes may be more strongly expressed in countries that have invested substantially in the public good of suppressing deaths from COVID-19 and, in particular, in cultures in which citizens hold moral expectations that their fellow citizens support the provisions of such goods.

Fig. 1: World map highlighting the countries included in study 1.



Countries are coloured by the share of vaccinated citizens in the population on the first day of data collection (December 2021 to January 2022).

Exclusion from family in 21 countries

Our initial examination focuses on cross-cultural exclusionary attitudes in the context of family relationships and, specifically, the level of antipathy if a close relative was marrying an unvaccinated (versus fully vaccinated) person. Such discriminatory attitudes in family relationships have been a key focus in previous cross-national research on prejudice along racial, ethnic and partisan lines[21,31]. Exclusion from family relationships is cross-culturally relevant, independently of the legal and democratic traditions of the country; discrimination on the basis of membership in politicized groups within families has also been shown to be highly disruptive for the families[32]; and, finally, relative to other forms of discriminatory attitudes (for example, support for state-sponsored discrimination), discrimination within families is something within the control of individuals and, therefore, something that can take immediate effect.

We used conjoint experiments in which respondents evaluated fictitious target profiles simultaneously randomized on six attributes, including their COVID-19 vaccination status. The conjoint experimental design yields causal traction, provides a cost-effective method for collecting large samples and enables us to examine a wide range of responses covering affective, cognitive and attitudinal components of prejudice[33].

Given our ambition to study discriminatory attitudes rather than generosity, we depart from previous work that relied on incentivized economic games (such as the dictator game)[8,9,18]. To help assess the validity of the conjoint experimental approach, we performed a number of tests. First, we show that people perceive that measures focusing on social interactions are more ecologically valid than those focusing on monetary transactions captured by economic games (paired sample t-test, $\Delta M = 0.15$, 95% confidence interval = 0.14–0.16, $t_{1,447} = 24.6$, $P < 0.001$; Extended Data Fig. 1 and Supplementary Information K), making the present findings less vulnerable to criticisms regarding ecological validity. Second, to assuage potential concerns about social desirability bias from self-reported measures, we

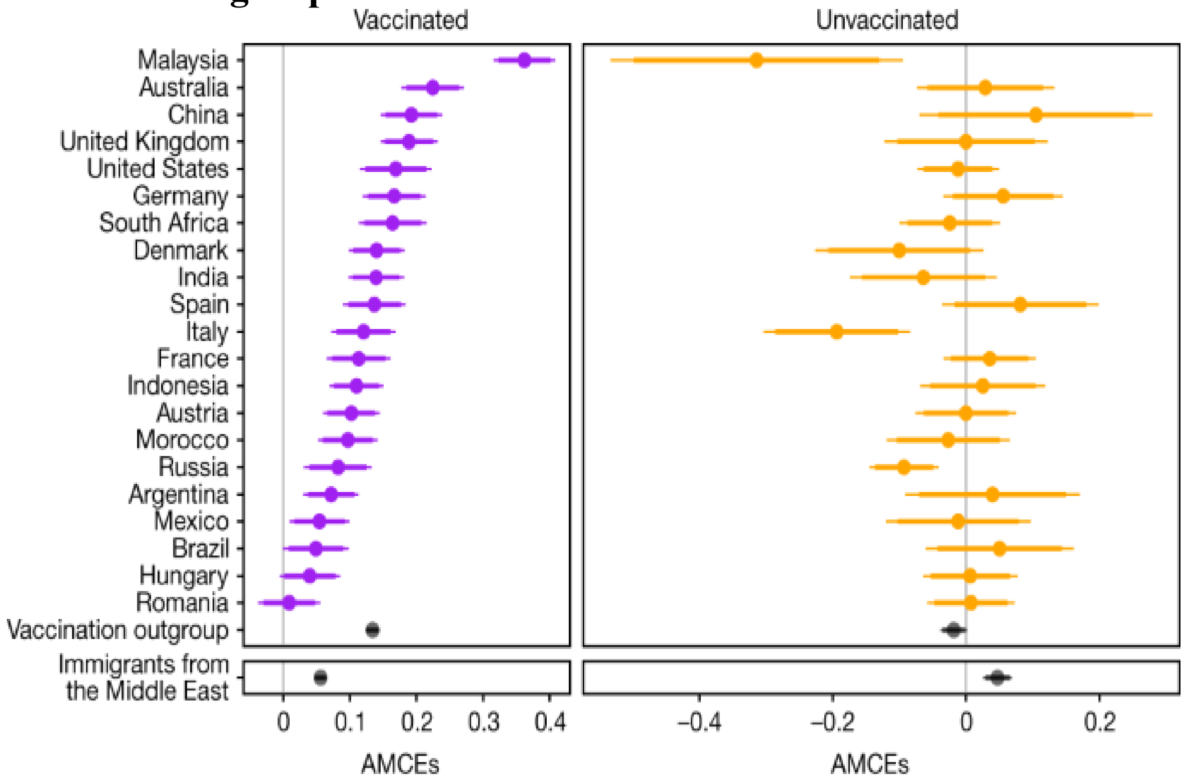
report experimental evidence indicating that people readily and openly admit their antipathy towards vaccination outgroups, even using a traditional, direct survey question ($M = 44\%$, 95% confidence interval = 0.40–0.48). Indeed, this estimate of antipathy is not statistically different from the estimate that we get using a forced-response technique, which was specifically designed to alleviate social desirability ($M = 39\%$, 95% confidence interval = 0.35–0.43, $\chi^2_{1,1,210} = 2.31$, $P = 0.13$; Extended Data Fig. 2 and Supplementary Information L). Finally, despite the presumed advantages of incentivized behavioural measures, we demonstrate that ingroup bias in generosity across vaccination groups is identical whether estimated with incentivized measures replicating previous research ($M = 29$; 95% confidence interval = 26–32, one-sample $t_{724} = 19.4$, $P < 0.001$) or with non-incentivized, self-reported measures ($M = 30$; 95% confidence interval = 28–33, one-sample $t_{722} = 21.0$, $P < 0.001$; $\Delta M = -1.45$, 95% confidence interval = −5.5–2.6, statistically equivalent to 0, TOST two-sample t-test, $t_{1,445} = 2.16$, $P < 0.05$; Supplementary Information M).

In the conjoint experiment for study 1, we adapted a widely used instrument of exclusionary reactions in family relationships[31] and examined a specific set of discriminatory attitudes: how unhappy would respondents be if a close relative was marrying an unvaccinated versus vaccinated person. Furthermore, we assessed the potential cognitive bases for discriminatory attitudes. First, we measured a reasonable basis for antipathy towards vaccination outgroups, namely, fear of infection[34] (note that, during the collection of these studies, the vaccine-evading Omicron variant was dominant[35], and vaccine-induced immunity against infection spread was waning[36] in most societies; this increased the chances of being infected by vaccinated people and therefore decreased the risk of interacting with unvaccinated individuals relative to vaccinated individuals). Although fear of infection is probably more pronounced among vaccinated individuals, some unvaccinated individuals have been found to hold the misinformed belief that vaccinated people themselves spread COVID-19 through vaccine shedding[37]. Second, we assessed the two key negative trait impressions underlying prejudice according to research on impression formation and prejudice—perceptions of untrustworthiness and unintelligence[29].

With the help of the YouGov and Ipsos survey agencies, we collected high-quality, quota-sampled, original survey data from 21 countries that had widespread access to vaccines against COVID-19 (study 1: 64,440 observations from 10,740 respondents). The data were collected in a diverse set of cultures from all inhabited continents of the world. As described in the 'Study 1' (Data and generalizability) section of the Methods, the samples can be considered to be representative of the countries' online populations (except for India). This large, cross-cultural dataset enables us to not only quantify discriminatory attitudes in a wide range of countries, but also investigate sources of cross-cultural variation in its levels. Note that our pre-registered analyses in study 1 focus on antipathy towards outgroups, pooling across respondent vaccination status. Given that we found large asymmetries by vaccination status, we report estimates separately for vaccinated and unvaccinated respondents. However, pooled estimates—which are reported in Supplementary Information F—mirror these results very closely given the relatively small share of unvaccinated respondents.

Our results reveal that vaccinated respondents ($n = 54,054$) exhibit exclusionary attitudes towards unvaccinated individuals (Fig. 2 (left)). On average, vaccinated respondents were 13 percentage points (average marginal component effects (AMCE) 95% confidence interval = 12–14, $z = 25.65$, $P < 0.001$) more unhappy when presented with an unvaccinated versus fully vaccinated target. Country-level estimates range from 1 to 36 percentage points. We can reject the null (at the 5% alpha level) in 19 out of the 21 countries. Malaysia is an outlier with very high exclusionary attitudes (AMCE = 36 percentage points, 95% confidence interval = 32–41, $z = 15.3$, $P < 0.001$), whereas the results in Romania (AMCE = 1 percentage point, 95% confidence interval = −4–6, $z = 0.39$, $P = 0.69$) and Hungary (AMCE = 4 percentage points, 95% confidence interval = −1–9, $z = 1.74$, $P = 0.08$) are inconclusive (further discussion is provided in Supplementary Information G). Interaction models estimating conditional AMCEs indicate that, although we observed exclusionary attitudes across all demographic groups, they are slightly stronger among highly educated (AMCE difference of 5 percentage points), female (by 4 percentage points), more affluent (by 3.5 percentage points) and older (by 2 percentage points) respondents (all $P < 0.05$; Supplementary Information D).

Fig. 2: Exclusionary attitudes in family relationships towards vaccination outgroups.



The average level of exclusionary attitudes in family relationships towards vaccination outgroups (that is, towards unvaccinated individuals for vaccinated respondents and towards vaccinated individuals for unvaccinated respondents; total $n = 64{,}440$). Exclusionary attitudes reflect being unhappy if a close relative married a person from the vaccination outgroup versus ingroup, with more positive coefficients indicating more exclusionary attitudes towards the outgroup relative to the ingroup. The purple and orange points denote country-level AMCE estimates ($n > 3{,}000$) for vaccinated and unvaccinated respondents, respectively. The black points denote the pooled sample and include an estimate for exclusionary attitudes towards immigrants from the Middle East. The error bars denote the 90% and 95% confidence intervals. For more details, see the 'Study 1' (Modelling) section of the Methods.

Meanwhile, unvaccinated respondents ($n = 10{,}386$) exhibit negligible exclusionary attitudes towards vaccinated individuals (Fig. 2 (right)). Their unhappiness is largely independent of the target's vaccination status, with an AMCE of only −2 percentage points (AMCE 95% confidence interval = −4–

0, $z = -1.81$, $P = 0.07$; the AMCE difference between vaccinated and unvaccinated respondents is 15 percentage points; 95% confidence interval = 13–18, $z = 13.33$, $P < 0.001$). Country-level estimates of exclusionary attitudes exhibited by unvaccinated respondents are noisy owing to the small sample sizes ($90 < n < 1{,}500$), ranging between −31 and 10 percentage points. Indeed, unvaccinated individuals in Malaysia, Italy and Russia even exhibit significant exclusionary reactions towards other unvaccinated individuals ($P < 0.01$), highlighting how anti-free-rider sentiments may take priority over sentiments related to ingroup favouritism[8].

To help assess the substantive size of these effects, it is helpful to compare them with exclusionary attitudes towards a group battling high levels of discrimination in many Western countries—immigrants from the Middle East[3]. Exclusionary attitudes towards unvaccinated individuals among vaccinated people (13 percentage points) is two and a half times greater than exclusionary attitudes towards Middle Eastern immigrants (5 percentage points, 95% confidence interval = 5–6, χ^2_1 ($n = 54{,}054$) = 23.83, $P < 0.001$). We do not suggest that the characteristics of these groups are comparable, but this finding nonetheless suggests that the substantive size of the exclusionary reactions facing unvaccinated individuals is high. Supplementary Figs. E.1 and E.2 juxtapose country-level estimates of exclusionary attitudes towards the two groups. Unvaccinated targets face significantly more exclusionary reactions compared with immigrants in 11 out of 21 countries, whereas immigrants do not face significantly more exclusionary reactions in any of the countries. Notably, exclusionary attitudes towards immigrants between vaccinated and unvaccinated individuals are substantively similar and not significantly different from 0 ($n = 64{,}440$, AMCE difference 1 percentage point, 95% confidence interval = −1–3, $z = 0.88$, $P = 0.38$), implying that asymmetry in the domain of vaccination cannot be easily explained by omitted variables or design effects (Supplementary Fig. E.3). Moreover, we did not find evidence that unvaccinated immigrants from the Middle East face disproportionate exclusionary attitudes compared with unvaccinated native individuals ($n = 64{,}440$, AMCE difference 1 percentage point, 95% confidence interval = 0–1, $z = 1.1$, $P = 0.27$; Supplementary Fig. E.4).

Stereotypes and exclusionary attitudes

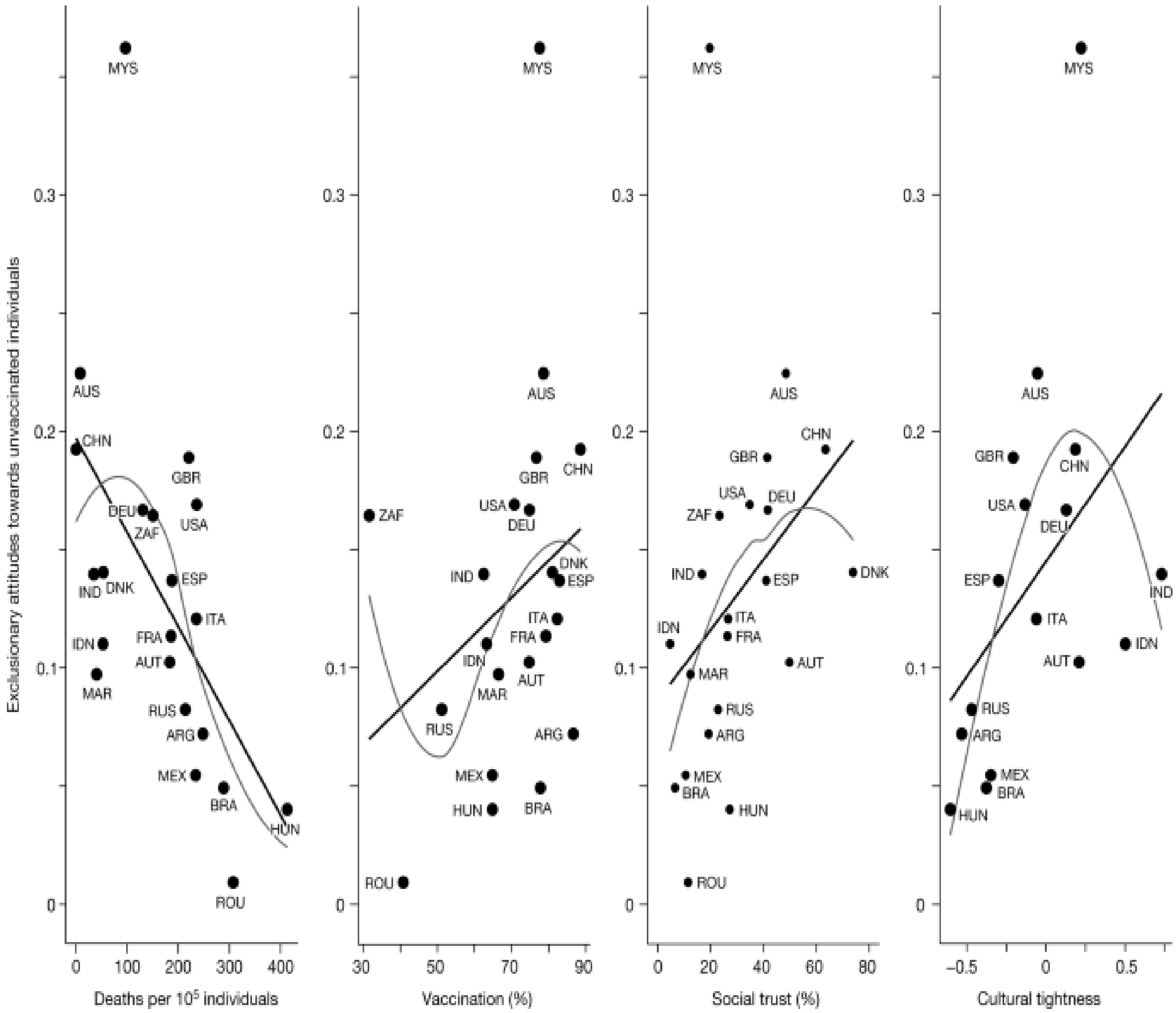
Next, we examined whether exclusionary attitudes reflect a heightened risk of infection or also activate more fundamental stereotypes. As displayed in Extended Data Fig. 3a, we find large experimental effects of vaccination status among vaccinated respondents on fear of infection (n = 54,054, 38 percentage points, 95% confidence interval = 37–40, z = 65.99, P < 0.001) and perceptions of untrustworthiness (13 percentage points, 95% confidence interval = 12–14, z = 27.36 P < 0.001). However, we also find an effect on incompetence (14 percentage points, 95% confidence interval = 13–15, z = 29.00, P < 0.001), suggesting that stereotypes of unvaccinated individuals extend beyond the domain of free-riding. As unvaccinated respondents (n = 10,386) exhibit insubstantial exclusionary reactions, it is not surprising that they also do not judge vaccinated respondents as untrustworthy (0 percentage points, 95% confidence interval = −2–2, z = 0.2, P = 0.84) or as incompetent (0 percentage points, 95% confidence interval = −2–2, z = 0.37, P = 0.71) either. If anything, they fear getting infected with SARS-CoV-2 by vaccinated people slightly less than by unvaccinated people (−2 percentage points, 95% confidence interval = −5–0, z = −2.1, P < 0.05). Country-level estimates of negative stereotypes against vaccination outgroups are shown in Supplementary Information C.

Our study also replicates a well-known finding from the impression-formation literature: impressions of trustworthiness have the greatest impact on overall exclusionary attitudes[29]. On the basis of a linear regression with respondent fixed effects, exclusionary attitudes are more closely associated with impressions of untrustworthiness (β = 0.24, 95% confidence interval = 0.23–0.25) than with impressions of incompetence (β = 0.17, 95% confidence interval = 0.16–0.18, Wald test for equal effects: χ^2_1 (n = 64,440) = 62.6, P < 0.001), or even infection concerns (β = 0.16, 95% confidence interval = 0.15–0.17, Wald test for equal effects: χ^2_1 (n = 64,440) = 112, P < 0.001; Extended Data Fig. 3b). Although concerns about infection risks do shape exclusionary attitudes towards unvaccinated individuals, these findings suggest that negative stereotypes further enhance these attitudes.

Culture and exclusionary attitudes

The results provide strong evidence that exclusionary attitudes against perceived free-riders in the domain of vaccinations emerge reliably across cultures, reflecting the deep-seated nature of the psychology of cooperation[6]. At the same time, it is clear that the strength of the observed exclusionary attitudes exhibits substantial cross-cultural variation (formal evidence is provided in Supplementary Information O.1). Figure 3 shows exclusionary attitudes towards unvaccinated individuals by vaccinated respondents against three pre-registered macro-indicators—COVID-19 deaths and vaccinations (both standardized to population size) and social trust—as well as the exploratory indicator cultural tightness. Whereas COVID-19 deaths and vaccination rates indicate society-wide investments in the public good of suppressing the epidemic, social trust (that is, the tendency to trust fellow citizens) and cultural tightness (that is, the strength of social norms and the degree of sanctioning within societies[38]) are indicators of the moral expectations of fellow citizens and the willingness to sanction violations of these expectations. Countries that managed to keep the number of deaths due to COVID-19 low show strong exclusionary attitudes towards unvaccinated individuals at around 20 percentage points on average. By contrast, countries that struggled to mitigate the epidemic show much lower exclusionary attitudes. The Spearman's rank-order correlation between death and prejudice is $\rho_{21} = -0.62$, 95% confidence interval = −0.83 to −0.26. At the same time, the association of exclusionary attitudes with actual vaccination levels is inconclusive $\rho_{21} = 0.38$ (95% confidence interval = −0.06–0.70). Although there is a tendency for nations with a high level of vaccination to display more exclusionary attitudes, and countries with lower compliance to display less, there are also considerable deviations from this trend, with outliers such as Argentina (high vaccination, low prejudice) and South Africa (high prejudice, low vaccination). In Supplementary Information G, we analyse policy stringency, which is a direct measure of how much national governments invested in suppressing infections. We find no reliable association between stringency and prejudice towards unvaccinated individuals ($\rho_{21} = 0.23$ 95% confidence interval = −0.22–0.6; Extended Data Fig. 4).

Fig. 3: The relationship between country-level indicators and cross-national levels of exclusionary attitudes among vaccinated individuals towards unvaccinated individuals.



The country-level indicators are country-level deaths from COVID-19; the national proportion of people expressing trust towards fellow citizens; the national proportion vaccinated against COVID-19; and cultural tightness scores. The labelled dots denote countries; the straight black lines denote best-fitting regression lines; and the grey curves denote loess curves. Spearman’s rank-order correlations across the four facets: deaths, ρ_{21} = −0.62, 95% confidence interval = −0.83 to −0.26; social trust, ρ_{21} = 0.57, 95% confidence interval = 0.19–0.81; vaccination, ρ_{21} = 0.38, 95% confidence interval = −0.06–0.70; tightness, ρ_{16} = 0.62, 95% confidence interval = 0.18–0.85. Total n = 64,440.

Previous research demonstrated that epidemic suppression hinges on citizens’ normative and moral expectations such that countries with higher social trust[39] and a tighter culture[40] suppressed the epidemic toll more

effectively. As shown in Fig. 3, these cultural differences are also associated with higher prejudice towards unvaccinated individuals. Specifically, exclusionary attitudes are higher in countries with higher social trust (Spearman's $\rho_{21} = 0.57$, 95% confidence interval = 0.19–0.81). In countries in which large majorities believe that 'most people can be trusted', the exclusionary reaction towards unvaccinated individuals is greater. Meanwhile, in countries in which most people believe that 'you need to be very careful in dealing with people', exclusionary attitudes are lower. Similarly, exclusionary attitudes are higher in countries with a tighter culture, oriented towards strong norms and the sanctioning of norm violations ($\rho_{16} = 0.62$, 95% confidence interval = 0.18–0.85). These latter associations suggest that cultures that place stronger moral expectations on individuals not only more effectively produce the public good of epidemic control[39,40] but also constitute a fertile ground for exclusionary attitudes against unvaccinated individuals, as they may be perceived to free-ride on the collective effort[8]. In Supplementary Information O.2, we provide robustness checks for these cross-cultural conclusions, addressing potential threats to the generalizability of data obtained through online surveys.

Antipathy across six countries

In study 2, we focused on the affective component of prejudice. Specifically, we conducted a pre-registered, conceptual replication of study 1 and, in the context of a conjoint experiment, asked the participants to rate fictitious individuals that vary in terms of vaccination status (as well as other attributes) on a seven-point like–dislike scale.

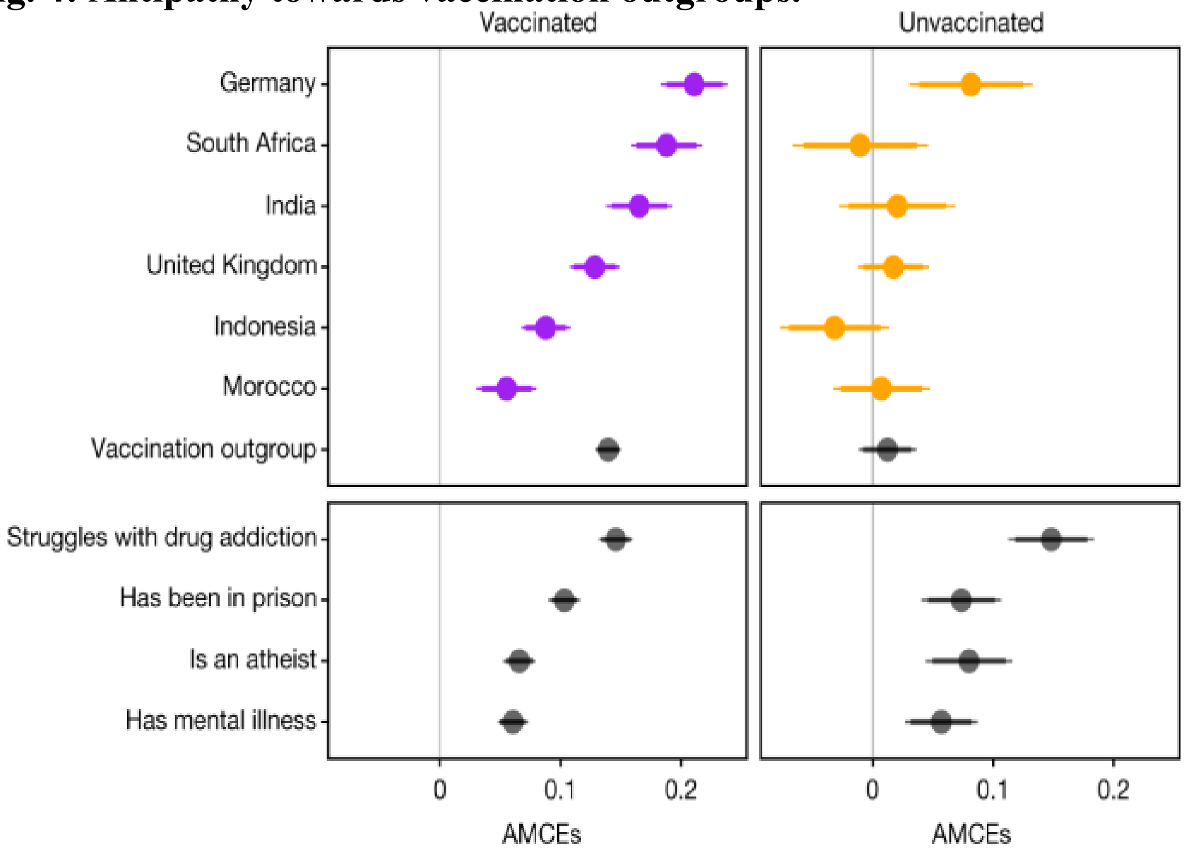
Study 1 also showed that exclusionary attitudes are intertwined with a fear of infection. Although fear of infection is a weaker correlate of exclusionary attitudes compared with trustworthiness impressions, the finding nonetheless raises the possibility that prejudice against unvaccinated individuals may be restricted to relationships characterized by physical interaction. The focus on pure antipathy in a neutral evaluation task enabled us to examine this possibility. Furthermore, to gain perspective on the level of antipathy across vaccination groups, in study 2, we also changed the benchmark group from Middle Eastern migrants to a more diverse set of four groups that are also frequent targets of prejudice: drug addicts, ex-convicts, people with mental

illnesses and atheists[4,5,41]. These groups were chosen to offer some variance on how much perceived control people have in determining their group membership and how much of a threat they pose on members of society.

With the help of the YouGov survey agency, study 2 was fielded in six countries (Germany, India, Indonesia, Morocco, South Africa and the UK), representing both Western affluent and non-Western developing nations. We recruited about 500 respondents per country, quota sampled on age, gender and region, as well as education in Germany and the UK (see Supplementary Information A). As before, each participant rated three pairs of target profiles (study 2: 3,045 respondents, 18,270 observations). All analyses, unless otherwise noted, were pre-registered (see Data Availability).

Our analyses show that vaccinated individuals feel antipathy towards unvaccinated individuals, even in a neutral evaluation task without any indication that the participants would physically meet the fictitious targets (Fig. 4). Across all six countries, we found that vaccinated respondents (n = 15,966) dislike unvaccinated targets more than vaccinated targets, on average by 14 percentage points (AMCE 95% confidence interval = 13–15, z = 25.94, P < 0.001). By contrast, unvaccinated respondents (n = 2,304) on average do not dislike vaccinated targets significantly more than unvaccinated targets (AMCE = 1 percentage points, 95% confidence interval = −1–4, z = 1.01, P = 0.31, although Germany is a significant outlier, AMCE = 8 percentage points, 95% confidence interval = 3–13, z = 3.12, P < 0.001). Note also that the substantive size of the antipathy expressed towards unvaccinated individuals remains high relative to the more diverse set of benchmarks. On average across the six countries, unvaccinated individuals are disliked as much as people who struggle with drug addiction (15 percentage points, 95% confidence interval = 13–16, Wald test for equal effects: χ^2_1 (n = 15,966) = 0.51, P = 0.47), and significantly more so than people who have been in prison (10 percentage points, 95% confidence interval = 9–11, χ^2_1 (n = 15,966) = 18.4, P < 0.001), who are atheists (7 percentage points, 95% confidence interval = 6–8, χ^2_1 (n = 15,966) = 67.5, P < 0.001) or who have a mental health illness (6 percentage points, 95% confidence interval = 5–7, χ^2_1 (n = 15,966) = 87.9, P < 0.001). Country-level estimates of antipathy towards each of the four benchmarks are provided in Supplementary Information I.

Fig. 4: Antipathy towards vaccination outgroups.



The average level of antipathy towards vaccination outgroups (that is, towards unvaccinated individuals for vaccinated respondents and towards vaccinated individuals for unvaccinated respondents; total $n = 18,270$). Antipathy reflects disliking a person from the vaccination outgroup versus the ingroup, with more positive coefficients indicating higher relative antipathy for the outgroup. The purple and orange points denote country-level AMCE estimates ($n > 3,000$) for vaccinated and unvaccinated respondents, respectively. The black points denote the pooled sample and include estimates for antipathy towards various other common targets of prejudice. The error bars denote the 90% and 95% confidence intervals. For more details, see the 'Study 2' (Modelling) section of the Methods.

Study 2 included an additional test. The finding from study 1 (that is, the widespread existence of exclusionary attitudes in personal relationships) may be less concerning if members of the groups of vaccinated and unvaccinated individuals are only weakly acquainted across group boundaries and if—consistent with intergroup contact theory—prejudice is

high only among individuals with less intergroup contact[42]. Study 2 therefore measured how many relatives and friends respondents have who belong to the vaccination outgroup. The analyses demonstrate that, although antipathy is indeed highest among people with no contact with the outgroups ($n = 18{,}270$, AMCE = 15 percentage points, 95% confidence interval = 13–16, $z = 20.36$, $P < 0.001$), it is substantial across all contact levels (AMCEs = 5–12 percentage points, $z \geq 2.4$, $P < 0.05$; Extended Data Fig. 5 and Supplementary Information J).

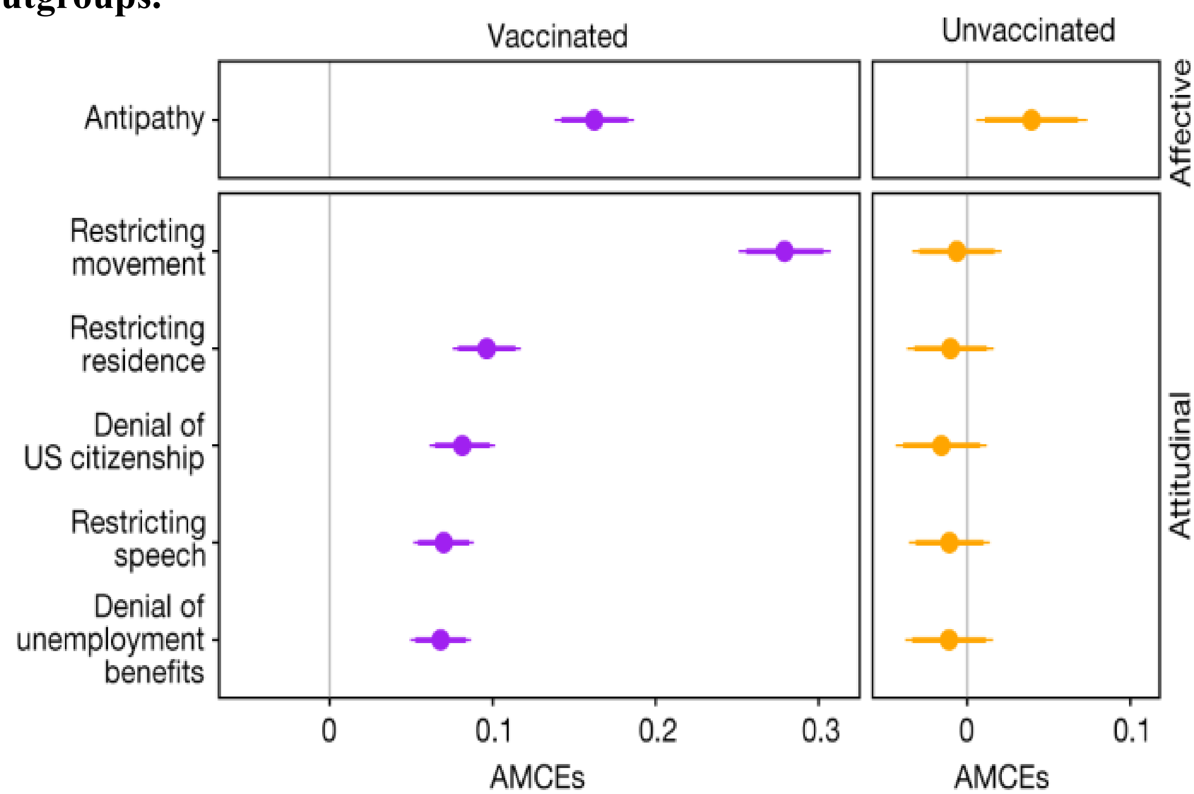
Restriction of rights in the USA

So far, the discriminatory attitudes we have investigated have been demonstrated only in the domain of private relationships. Thus, study 3 examined whether discriminatory attitudes extend into the domain of publicly recognized rights. As the recognition of such rights differs across cultures, study 3 was conducted in the USA—a country with historical recognition of fundamental rights and freedoms[43]. Study 3 was identical to study 2, except that the study also included five new outcomes in addition to the measure of antipathy, and the answers were obtained on binary scales. Specifically, respondents were asked to evaluate the target's freedom of movement ("This person should be allowed to sit next to me in public transportation"), freedom of residence ("This person should be allowed to move into my neighbourhood"), freedom of speech ("This person should be allowed to express their political views on social media freely, without fear of censorship"), access to citizenship ("This person should receive US citizenship, if they are eligible and apply for it") and access to unemployment benefits ("This person should receive unemployment benefits, if they are eligible and apply for it"). We collected data through YouGov from 1,448 US Americans quota sampled on age, gender, region, education and race. Each respondent evaluated five pairs of targets yielding a final sample size of 14,480 observations. All analyses, unless otherwise noted, were pre-registered (see Data availability). The survey also included the methodological studies discussed in relation to study 1, which are reported in detail in Supplementary Information K–M.

The results (Fig. 5) demonstrate that exclusionary attitudes are not restricted to the domain of private relationships. Vaccinated Americans not only feel

greater antipathy towards unvaccinated Americans by 16 percentage points (95% confidence interval = 14–19, $z = 13.09$, $P < 0.001$), but they are also 28 percentage points less likely to respect their freedom of movement (95% confidence interval = 25–31, $z = 19.4$, $P < 0.001$), 10 percentage points less likely to respect their freedom of residence (95% confidence interval = 8–12, $z = 9.1$, $P < 0.001$), 8 percentage points less likely to support their application for citizenship (95% confidence interval = 6–10, $z = 7.98$, $P < 0.001$) and 7 percentage points less likely both to respect their freedom of speech and to support their applications for welfare benefits (95% confidence interval = 5–9, $z = 7.23$ and 7.44, respectively, $P < 0.001$). Vaccinated respondents expressed significantly higher exclusionary attitudes towards unvaccinated individuals than against atheists on all six outcomes, against mentally ill on five outcomes, and against individuals who have been in prison or struggle with drug addiction on three outcomes. By contrast, they do not express significantly higher exclusionary attitudes towards any of the benchmark groups on any of the outcomes compared with unvaccinated individuals (details on all Wald tests are provided in Supplementary Information I).

Fig. 5: Affective and attitudinal prejudice against vaccination outgroups.



Affective and attitudinal prejudice against vaccination outgroups in the USA (that is, towards unvaccinated individuals for vaccinated respondents and towards vaccinated individuals for unvaccinated respondents; $n = 14{,}480$). Prejudice reflects relative antipathy towards and support for restricting the rights and freedoms of the outgroup relative to the ingroup. More positive coefficients indicate higher prejudice. The purple and orange points denote AMCEs among vaccinated and unvaccinated respondents, respectively. The error bars denote the 90% and 95% confidence intervals. For more details, see the 'Study 3' (Modelling) section of the Methods.

The results of study 3 also indicate that unvaccinated Americans also have some negative sentiment towards vaccinated individuals (4 percentage points, 95% confidence interval = 1–7), but unvaccinated Americans are neither more nor less likely to restrict their rights or freedoms. Finally, additional analyses (Supplementary Information N) indicate that the antipathy of vaccinated Americans towards unvaccinated individuals is predictive of their support for restricting the rights of unvaccinated individuals (Spearman's rank order correlations $0.35 < \rho_{1,448} < 0.44$).

Discussion

Research on political polarization warns that, if sociopolitical disagreement—even if based on legitimate grievances—permeates interactions between citizens, it can contribute to the entrenchment of conflict[21]. Here we show that individuals who are vaccinated against COVID-19 express negative attitudes against unvaccinated individuals in the form of antipathy, stereotypes, support for exclusion from family relationships and support for removal of political rights. In total, these four forms of discriminatory attitudes are consistent with the observation of prejudice according to standard definitions in social psychology. We examined and obtained evidence in support of all four reactions in the USA. In the other countries, we examined only some but not all forms of discriminatory attitudes and found evidence in support of the specific negative reactions examined. The only exceptions were Hungary and Romania, in which we did not find evidence in support of discriminatory attitudes. Furthermore, we found that discriminatory attitudes towards unvaccinated individuals is as high or higher than discriminatory attitudes directed towards other common and

diverse targets of prejudice including immigrants, drug-addicts and ex-convicts. At the same time, the results demonstrate that prejudice is mostly one-sided. Only in the USA and Germany did we find that unvaccinated individuals feel some antipathy towards vaccinated individuals but even here we did not find statistical evidence in favour of negative stereotyping or exclusionary attitudes.

The finding that vaccinated individuals are prejudiced against unvaccinated individuals but that there is no evidence for the reverse is consistent with studies on the psychology of cooperation[6,7] and previous research on vaccinations. The cue that someone refuses to take up a vaccine activates psychological mechanisms designed to deter perceived free-riders among vaccinated individuals[8,9]. Consistent with the deep-seated nature of anti-free-rider sentiments, the observation of substantial and culturally widespread discriminatory attitudes, including support for denial of fundamental rights, suggests that negative reactions are easily triggered in the context of perceived public goods. At the same time, the results also reveal that some cultures are especially prone to react with prejudice. Consistent with an anti-free-rider perspective, vaccinated individuals in cultures with stronger cooperative norms react more negatively against unvaccinated individuals. Such norms are more reliably associated with cross-cultural differences in discriminatory attitudes than actual country-level differences in government efforts to produce epidemic control. What seems to trigger discriminatory attitudes towards unvaccinated individuals is less governments' efforts to reduce deaths from COVID-19 and more how such efforts resonate with larger cultural norms and perceived obligations.

In this regard, note that the decision to refuse vaccination against COVID-19 may reflect many factors beyond a moral failure to appreciate collective goals. A recent review of almost 100 empirical studies identified 18 robust correlates of COVID-19 vaccine hesitancy in high-income countries[44]. Even if negative stereotypes are statistically true, they are unlikely to adequately capture the full motivations of every individual. For example, an unvaccinated person may have medical conditions[45], immunity from previous infection[46], a history of mental health issues that may intensify fear of vaccinations[47], negative past experiences with health authorities (especially as members of a minority group)[48], concerns due to country-

specific public health scandals[49] or ethical considerations about vaccine equity[50].

Although moralistic communication of collective responsibilities may be an effective strategy to increase vaccination uptake[8], such strategies may have unintended negative consequences in the form of eliciting prejudice[51], especially in cultures with strong cooperative norms. Research on prejudice towards minority groups warns that experiences of prejudice and discrimination may have negative long-term effects, hurting well-being[52], eroding identification with majority society[53] and driving mistrust of the state, including health authorities[54]. If the consequences of prejudice towards unvaccinated individuals resemble the consequences of prejudice against minority groups, they may exacerbate the mistrust and alienation that led to vaccine refusal in the first place[13,20].

In the short run, prejudice towards unvaccinated individuals may complicate pandemic management. In the long run, it may mean that societies leave the pandemic more divided than they entered it. Finally, our findings also offer a lesson for global challenges beyond the current pandemic. Large social crises—for example, the climate crisis—are often characterized by collective action dilemmas due to the need for substantial behaviour change among the public[55]. To effectively manage such crises, the authorities should seek to avoid fuelling deep animosity between citizens. Indeed, as moral condemnation is often easily and spontaneously activated among the public during a crisis[11], the authorities and politicians should consider tempering social animosities as an important part of their mandate, especially when societal conflict becomes more entrenched.

Methods

Study 1

Data and generalizability

We collected data between 3 December 2021 and 28 January 2022 from the following 21 countries: Argentina, Australia, Austria, Brazil, China,

Denmark, France, Germany, Hungary, India, Indonesia, Italy, Malaysia, Mexico, Morocco, Romania, Russia, South Africa, Spain, the UK and the USA. Data were collected through online panels by Ipsos in China, and by YouGov in all other countries. All of the participants provided informed consent and were reimbursed according to their standing agreements with the data provider. All of the studies (studies 1–3) were exempt from formal ethical review by Danish law. As per section 14(2) of the Act on Research Ethics Review of Health Research Projects, "notification of questionnaire surveys ... to the system of research ethics committee system is only required if the project involves human biological material." The studies fully comply with Aarhus University's Code of Conduct and with the ethical standards set by the Danish Code of Conduct for Research Integrity. All hypotheses, materials and analyses were pre-registered at the OSF (https://osf.io/6teug).

We sought to recruit 500 adult respondents from each country, quota sampling on age, gender and region of residence, and—conditional on feasibility—also education (in Australia, Brazil, Denmark, France, Germany, Italy, Mexico, Russia, Spain, the UK and the USA) and race (in the USA). Demographic details are provided in Extended Data Table 1. Quotas were always set to mimic the national population, except in Indonesia, Morocco and Malaysia, where, due to feasibility issues, they were set to the demographic characteristics of the online population and, in India, where they were set to the demographic characteristics of the national urban population. Respondents who failed a simple bot test were screened out at the beginning of the survey. For additional demographic information, as well as more details on the quotas set, see Supplementary Information A. Questionnaires were translated to the official language of the country by professional translators employed by a translation agency (deviations from this rule are described in Supplementary Information B). Each translation was independently quality-checked by another translator at the agency, and a native speaker recruited by the researchers.

Our samples cover a diverse set of cultures from all inhabited continents of the world. That said, our sample intentionally excludes low-income countries in which COVID-19 vaccines were not yet widely available to the public and where, accordingly, we would not expect vaccination status to lead to prejudice. Furthermore, despite the quotas set, our samples are not

fully nationally representative as they exclude members of society who have no internet access, or face other systemic disadvantages (for example, are illiterate or do not speak the official language of the country). Previous research concludes that, for results from cross-national data collected through YouGov's online panels in low- and middle-income countries, "it is better to think of [them] as representative of the online population."[56]. At the same time, during a pandemic, online surveys constitute a safe and efficient data collection method that enables voices from diverse cultures to be heard. Consistent with this, the World Health Organization refers to the use of online surveys as the standard approach for behavioural insights during the pandemic[57]. Our main survey vendor, YouGov, enabled Imperial College London to capture global behavioural dynamics during the pandemic (https://www.imperial.ac.uk/global-health-innovation/what-we-do/our-response-to-covid-19/covid-19-behaviour-tracker/).

Previous research suggests that cross-cultural differences can be reliably studied using online surveys[58,59,60] and that studies using experimental designs (as we do) are particularly robust across a variety of sampling methods[61,62]. Nonetheless, it is important to acknowledge and address limitations to the generalizability of online surveys. The key threat in this regard is whether differences between online and national populations may endanger the robustness of the cross-cultural conclusions. To examine the consequences of this threat to inference, we report multiple robustness checks of the data and analyses. First, Supplementary Information A compares the most relevant objective benchmark—actual vaccinations against COVID-19 in the adult national population, against those observed in the surveys, finding high correspondence. Second, Supplementary Information D directly examines treatment heterogeneity in two of the key indicators associated with internet access—education and income—and finds very little treatment heterogeneity, even in low- and middle-income countries (Extended Data Fig. 6). Third, Supplementary Information O.2 directly quantifies the potential threat to inference and examines the robustness of the cross-cultural conclusions to potential differences in prejudice between online and offline populations. It finds that the cross-cultural conclusions are robust to even the extreme assumption that the offline population holds zero prejudice against unvaccinated individuals (Extended Data Fig. 7). Fourth, Supplementary Information O.2 also reports

a stress test that examines the consequences of simultaneous violations of our two conjectures that (1) our samples represent the online populations and (2) that the represented and non-represented populations show similar prejudice. This test finds that the conclusion of cross-culturally pervasive prejudice would hold even if both of our conjectures were wrong. Overall, both previous research and extensive robustness analyses strongly suggest that our conclusions hold as stated.

Experimental design

Our design is a subtle, conjoint experimental implementation of E. S. Bogardus' classic social distance scale[3] (that said, this does not mean that all of the countries in our sample must be prejudiced against Middle Eastern immigrants; although it has been important for us that none of our samples come from the Middle East (as it would render the attribute meaningless), factors like a shared religious identity may temper prejudice against Middle Easterners; to minimize this latter effect, we deliberately avoided using the term 'Muslim'). We presented participants with brief descriptions of a series of fictitious individuals and asked them to imagine that these are people whom one of their close relatives intends to marry. One of the six attributes describing these target individuals has been their COVID-19 vaccination status, randomly varying between 'fully vaccinated' and 'unvaccinated'. Importantly, this is a minimalist manipulation of vaccination status, simply labelling target individuals with their group membership and therefore offering no reason or justification for their choice.

We were interested in whether participants would have higher exclusionary attitudes against unvaccinated individuals marrying into their families. To benchmark the magnitude of this hypothesized prejudice, another attribute, labelled family background, distinguished between people who were "born and raised in [the respondent's country]" and people who "immigrated from the Middle East." Middle Eastern immigrants are an excellent benchmark, as widespread prejudice against them has been widely documented[3]. The other four attributes (age, occupation, hobbies and personality) were included to increase ecological validity and to reduce experimenter demand and social desirability. All in all, we collected data from 10,740 individuals. We report in the pre-registration a sensitivity analysis justifying our sample size.

Extended Data Table 2 displays the six attributes and their levels. Each target profile was sampled completely at random from the 2 × 2 × 6 × 6 × 6 × 5 = 4,320 unique combinations of attribute levels. Following best practices in the literature[33], we presented two targets at a time, side by side. Each respondent rated six random targets independently across three trials, yielding a total sample size of 64,440 observations. We also randomized between respondents the order in which the attributes appeared.

Measures

Respondents rated each target profile independently by indicating their agreement or disagreement with a series of four statements on a simple yes/no scale. Specifically, we measured respondents' exclusionary attitudes with the statement "I would be unhappy if this person married one of my close relatives"; fear of infection with the statement "I would be afraid that this person infected me or my family with COVID-19"; perceptions of intelligence with "I think this person is unintelligent"; and trustworthiness with "I think this person is untrustworthy."

We also collected background information on all of the respondents. Importantly, before the treatment, we asked whether respondents themselves were vaccinated or not. We labelled all respondents who received at least one vaccine as "vaccinated," and all other respondents, including those who refused to answer the question, as "unvaccinated". We also rely on demographic data shared by the survey provider, which we dichotomized into male and female respondents, older and younger respondents (by splitting at the sample median in each country), respondents with and without a completed higher (tertiary) education and, finally, low-income respondents with a gross household income of less than 70% of the national median and respondents above that threshold.

Finally, our analyses rely on a series of country-level predictors. We measure pandemic severity with the cumulative number of confirmed COVID-19 deaths per 100,000 people in the total population on the first day of data collection in the country as measured by Johns Hopkins University. We measure the vaccination rate of the country with the total number of people who received at least one vaccine dose per 100 people in the total population

on the first day of data collection in the country as measured by Our World in Data. We measure social trust with the proportion of respondents who said "most people can be trusted" (versus "you need to be very careful in dealing with people") in the latest World Values Survey data available for the country. Finally, we use cultural tightness–looseness scores from ref. [38] as a post hoc predictor of prejudice against unvaccinated individuals. Note that tightness scores were available for 16 out of our 21 countries, and Denmark, France, Morocco, Romania and South Africa are omitted from these analyses.

Modelling

According to standard practices in the literature on conjoint experiments[63], we analysed our data using ordinary least squares regression models regressing one by one the four outcomes on the six categorical attributes. The models include post-stratification weights. We cluster standard errors on respondents. Our four hypotheses were evaluated on the AMCE of vaccination attribute (scaled to indicate outgroups) on the four outcomes. AMCEs can be interpreted as the percentage point change in the proportion of respondents reporting exclusionary attitudes, perceived untrustworthiness, unintelligence or fear of infection, caused by changing a target's vaccination status from ingroup to outgroup. All significance tests reported in the manuscript—unless otherwise noted—were two-sided.

To estimate country-level effects, we rerun these models in each of the 21 countries separately. To estimate heterogeneities in prejudice between vaccinated and unvaccinated respondents, as well as between demographic groups, we rerun models on split samples. To estimate country-level relationships between exclusionary attitudes and macro-level indicators, we relied on descriptive plots and Spearman's rank order correlations.

Our identification strategy rests on the random assignment of vaccination status to target individuals[63]. We report the standard diagnostic tests for conjoint experiments in Supplementary Information O.6. We find little reason for concern, although we acknowledge that participants speeding through the experiment dilute the observed experimental effects (Supplementary Fig. O.21) and that there are some carry-over effects for

exclusionary attitudes but not the other three outcomes (Supplementary Fig. O.24). We also note that insofar as some of our respondents falsely claim to be vaccinated, our estimates of prejudice towards unvaccinated individuals are likely to be too conservative.

Finally, in Supplementary Information O, we report robustness tests. All our conclusions replicate when our regressions are implemented in a Bayesian multilevel framework, when we drop post-stratification weights and when we excluded respondents who claimed that they had no previous opportunity to get vaccinated.

Study 2

Study 2 sought to conceptually replicate and extend the results of study 1. First, it relies on an alternative, purely affective measure of prejudice, which cannot be confounded by concerns of infection risk. Second, it uses an alternative set of benchmark groups to get additional perspective on the substantive size of the prejudice faced by unvaccinated individuals. Third, it tests whether antipathy against vaccination outgroups is lower among people who have more contact with members of the outgroup. Finally, it conceptually replicates our findings in a period after the Omicron wave had receded and when concerns about the pandemic were less outspoken.

Data and design

Our data were collected in May 2022 from six countries: Germany, India, Indonesia, Morocco, South Africa and the UK. As before, our data provider, YouGov, quota sampled a minimum of 500 respondents per country from online panels. All of the participants provided informed consent and were reimbursed according to their standing agreements with the data provider. The study was exempt from formal ethical review (see the ‘Study 1’ section above).

The design of study 2 closely mirrors the conjoint experimental design described above for study 1. For the sake of brevity, we therefore focus on deviations here. First, we omitted the framing of the relationship between respondent and target as a prospective close family member. Instead, we

simply presented target individuals to respondents, whom they were asked to evaluate on a standard like–dislike scale. Second, we replaced the family background attribute with a new attribute called personal information. Under this inconspicuous label, we included references to membership in one of four groups, which are well documented for facing various levels of prejudice: drug addicts, mentally ill, convicts and atheists. As a neutral comparison, the attribute also had a control condition: “no additional information”. These four groups intentionally vary in the extent that membership is conditional on personal choice versus luck, and whether they pose a danger on others. A detailed description of all of the attributes (including two minor changes on background attributes) is provided in Supplementary Information B. Finally, beyond personal vaccination status, we also measured the personal experience of contact with vaccination outgroups and modelled the antipathy towards outgroups conditional on this variable.

As before, the target profiles were sampled completely at random from the $2 \times 5 \times 6 \times 5 \times 6 \times 5 = 9{,}000$ unique combinations of attribute levels (Extended Data Table 3. Each respondent evaluated three pairs of targets. This yields a final sample size is 18,270 observations from 3,045 individuals. All hypotheses, materials and analyses were pre-registered at the OSF (https://osf.io/a7hsu).

Measures

To measure contact with vaccination outgroup members, we asked, before treatment, “how many relatives and friends do [respondents] have who are [not] vaccinated against COVID-19?” The response options were: None at all, 1–2, 3–5, 6–10 and more than 10. The question always referred to the outgroup, based on a measure and categorization of personal vaccination status identical to the one used in study 1.

General impressions of the targets were measured on a standard seven-point Likert scale from strongly dislike to strongly like. Respondents were prompted for each target to indicate “how much do [they] like or dislike person [A–F]”.

Modelling

We followed the same modelling strategy as in study 1. We recoded the continuous dependent variables to the 0–1 range, with higher values indicating more dislike. To investigate whether the respondents with more contact with outgroups express lower antipathy towards them, we pre-registered an interaction model, estimating antipathy conditional on contact levels, treated as a categorical variable with no contact as the reference category.

Study 3

The primary ambition of study 3 was (1) to extend previous results relying on a wider range of outcome measures tapping into various forms of prejudice. Moreover, it also included two additional experiments, relying on alternative paradigms for measuring prejudice and generosity across groups. Accordingly, (2) we tested whether people are less generous with unvaccinated individuals in an economic game both with and without monetary incentives[8,9,18]. (3) We also measured prejudice using the Bogardus' family context, both with a standard direct question and using a forced response technique[64]. Relying on these data, (4) we can investigate whether social desirability biases the propensity to admit prejudice towards vaccination outgroups. Finally, (5) we collect data to understand whether social interactions as a context for studying negative attitudes across vaccination outgroups are less vulnerable to criticism regarding ecological validity than standard economic games. All hypotheses, materials and analyses were pre-registered at the OSF (https://osf.io/ypc6a).

Data and design

We collected data from 1,448 adults living in the USA in May 2022 (simultaneously with study 2). As before, respondents were recruited from YouGov's online panel using quota sampling on gender, age, region, education and race. All of the participants provided informed consent and were reimbursed according to their standing agreements with the data provider. The study was exempt from formal ethical review (see the 'Study 1' section above).

The conjoint experimental design was identical to that of study 2, except each of the participants rated five pairs of target profiles. This yielded a final sample size of 14,480 observations in the conjoint experiment. The study also included two additional experiments. First, replicating ref. [18], the respondents participated in a dictator game, in which an allocator can give some of their 100 points endowment to another player, the recipient. All of the participants played in the role of the allocator, and were randomly matched with another respondent in the survey (post hoc), about whom they only knew whether they are vaccinated or unvaccinated against SARS-CoV-2. We used the strategy method and elicited an allocation for both types of partners (in a random order). We experimentally manipulated between participants whether the participants played for a monetary incentive. Specifically, we informed a random half of the participants that the points they divide in the game are worth money at a rate of 100 points = 250 YouGov points. We calibrated this to correspond to roughly $0.20, an incentive equal[8,9] or higher[18] than those used in previous research.

Second, we also measured whether respondents agreed or disagreed with the statement "I would be unhappy if a person [not] vaccinated against COVID-19 married one of my close relatives." The statement always referred to vaccination outgroups. Importantly, we manipulated between participants whether the question was asked directly or embedded in a forced response design, which uses a randomization device to mask the responses of individual respondents, while retaining the ability to estimate the sample-level agreement. Specifically, using a third-party random-number generator, respondents drew an integer between 1 and 6. If they got 1 or 6, they were 'forced' to respond agree or disagree, respectively. If they got anything in between, they answered freely, according to their true preference. This method is designed to remove social desirability bias from sensitive survey questions[64].

Measures

In the conjoint experiment, the participants evaluated six statements for each target, indicating whether they (1) like the target, whether they support the target's applications for (2) citizenship and (3) unemployment benefits;

and if they respect the target's (4) freedom of expression, (5) freedom of residence and (6) freedom of movement.

After a brief explanation of the rules of the dictator game (referred to as an allocation task), the participants were asked "How many points would [they] give to this vaccinated/unvaccinated person?" We calculated the difference in points allocated to vaccination outgroups versus ingroups. Higher scores indicate more ingroup favouritism.

In the third experiment, we calculated the proportion of participants who indicated that they would be unhappy if someone from the vaccination outgroup would marry into their family. This is a simple proportion of 'yes' answers to the direct question but, in the forced-response condition, we must correct the counts to account for the fact that a third of all respondents are forced to respond one way or another. Accordingly, we subtract 1/6 of the total sample size both from the agree and the disagree responses. We test whether the proportion of prejudiced respondents is statistically significant from 0 and whether prejudice is higher or lower in the forced response condition, compared with the direct question condition.

For the measure on the best context to study discriminatory attitudes against vaccination outgroups, we operationalize ecological validity as the frequency with which people encounter situations that are similar to the one described in the study. Specifically, the participants answered how often or rarely they encounter six situations, three of which describe social interactions (for example, "I get upset when I think about interacting with all the people [not] vaccinated against COVID-19.") and three describing monetary transactions (for example, "I consider donating money to individuals [not] vaccinated against COVID-19.").

Modelling

For the conjoint experiment, we followed the very same analysis strategy as described for study 1. For the dictator game, we conducted simple *t*-tests to estimate whether the participants were significantly more generous towards their in-group members, and whether the size of this in-group favouritism was affected by the incentives offered. For Bogardus' measure of social

distance, we estimated the uncertainty of the estimates both using standard CIs and also using a χ^2 test to determine whether social desirability biases estimates compared with the direct question.

Finally, the measures of ecological validity were compared using *t*-tests.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All pre-registrations, data and materials necessary to reproduce or replicate our analyses are available online (https://osf.io/7hszd).

Code availability

All computer code necessary to reproduce or replicate our analyses is available online (https://osf.io/7hszd).

References

1. Troiano, G. & Nardi, A. Vaccine hesitancy in the era of COVID-19. *Publ. Health* **194**, 245–251 (2021).

2. Karafillakis, E., Van Damme, P., Hendrickx, G. & Larson, H. J. COVID-19 in Europe: new challenges for addressing vaccine hesitancy. *Lancet* **399**, 699–701 (2022).

3. Bansak, K., Hainmueller, J. & Hangartner, D. How economic, humanitarian, and religious concerns shape european attitudes toward asylum seekers. *Science* **354**, 217–222 (2016).

4. Gervais, W. M. et al. Global evidence of extreme intuitive moral prejudice against atheists. *Nat. Hum. Behav.* **1**, 0151 (2017).

5. Pescosolido, B. A., Medina, T. R., Martin, J. K. & Long, J. S. The "backbone" of stigma: identifying the global core of public prejudice associated with mental illness. *Am. J. Publ. Health* **103**, 853–860 (2013).

6. Price, M. E., Cosmides, L. & Tooby, J. Punitive sentiment as an anti-free rider psychological device. *Evol. Hum. Behav.* **23**, 203–231 (2002).

7. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).

8. Korn, L., Böhm, R., Meier, N. W. & Betsch, C. Vaccination as a social contract. *Proc. Natl Acad. Sci. USA* **117**, 14890–14899 (2020).

9. Weisel, O. Vaccination as a social contract: the case of COVID-19 and US political partisanship. *Proc. Natl Acad. Sci. USA* **118**, e2026745118 (2021).

10. Jørgensen, F., Bor, A. & Petersen, M. B. Increased pressure leads to decreased trust among the unvaccinated: effects of the announcement of the reintroduction of COVID passports in Denmark. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/j49zg (2021).

11. Bor, A., Jørgensen, F., Lindholt, M. F. & Petersen, M. B. Moralizing the COVID-19 pandemic: self-interest predicts moral condemnation of other's compliance, distancing and vaccination. *Polit. Psychol.* https://doi.org/10.1111/pops.12835 (2022).

12. Kashte, S., Gulbake, A., El-Amin III, S. F. & Gupta, A. COVID-19 vaccines: rapid development, implications, challenges and future prospects. *Hum. Cell* **34**, 711–733 (2021).

13. Lindholt, M. F., Jørgensen, F., Bor, A. & Petersen, M. B. Public acceptance of COVID-19 vaccines: cross-national evidence on levels and individual-level predictors using observational data. *BMJ Open* **11**, e048172 (2021).

14. Jørgensen, F., Bor, A., Rasmussen, M. S., Lindholt, M. F. & Petersen, M. B. Pandemic fatigue fueled political discontent during the COVID-

19 pandemic. *Proc. Natl Acad. Sci. USA* **119**, e2201266119 (2022).

15. Bor, A., Jørgensen, F. & Petersen, M. B. The COVID-19 pandemic eroded system support but not social solidarity. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/qjmct (2021).

16. Lilleholt, L., Zettler, I., Betsch, C. & Böhm, R. Pandemic fatigue: measurement, correlates, and consequences. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/2xvbr (2020).

17. Schuessler, J., Dinesen, P. T., Østergaard, S. D. & Sønderskov, K. M. Public support for unequal treatment of unvaccinated citizens: evidence from Denmark. *Soc. Sci. Med.* **305**, 115101 (2022).

18. Henkel, L., Sprengholz, P., Korn, L., Betsch, C. & Böhm, R. The association between vaccination status identification and societal polarization. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-022-01469-6 (2022).

19. Rosenfeld, D. L. & Tomiyama, A. J. Jab my arm, not my morality: perceived moral reproach as a barrier to COVID-19 vaccine uptake. *Soc. Sci. Med.* **294**, 114699 (2022).

20. Petersen, M. B., Bor, A., Jørgensen, F. & Lindholt, M. F. Transparent communication about negative features of covid-19 vaccines decreases acceptance but increases trust. *Proc. Natl Acad. Sci. USA* **118**, e2024597118 (2021).

21. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the United States. *Ann. Rev. Polit. Sci.* **22**, 129–146 (2019).

22. Crandall, C. S. & Eshleman, A. A justification-suppression model of the expression and experience of prejudice. *Psychol. Bull.* **129**, 414–446 (2003).

23. Dixon, J., Levine, M., Reicher, S. & Durrheim, K. Beyond prejudice: are negative evaluations the problem and is getting us to like one another more the solution? *Behav. Brain Sci.* **35**, 411–425 (2012).

24. Gibson, J. L., Claassen, C. & Barceló, J. in *At the Forefront of Political Psychology: Essays in Honor of John L. Sullivan* (eds Borgida, E. et al.) 55–78 (Routledge, 2020).

25. Stangor, C., Sullivan, L. A. & Ford, T. E. Affective and cognitive determinants of prejudice. *Soc. Cogn.* **9**, 359–380 (1991).

26. Cosmides, L. & Tooby, J. in *The Handbook of Evolutionary Psychology* (ed. Buss, D. M.) 584–627 (Wiley, 2015).

27. Ahlskog, R. *Essays on the Collective Action Dilemma of Vaccination*. PhD thesis, Acta Universitatis Upsaliensis (2017).

28. Schauer, F. *Profiles, Probabilities, and Stereotypes* (Harvard Univ. Press, 2006).

29. Fiske, S. T., Cuddy, A. J. C. & Glick, P. Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* **11**, 77–83 (2007).

30. Neely, S. R., Eldredge, C., Ersing, R. & Remington, C. Vaccine hesitancy and exposure to misinformation: a survey analysis. *J. Gen. Intern. Med.* **37**, 179–187 (2022).

31. Bogardus, E. S. Measurement of personal-group relations. *Sociometry* **10**, 306–311 (1947).

32. Hernandez, R. A., & Colaner, C. "This is not the hill to die on. Even if we literally could die on this hill": examining communication ecologies of uncertainty and family communication about COVID-19. *Am. Behav. Sci.* **65**, 956–975 (2021).

33. Hainmueller, J., Hangartner, D. & Yamamoto, T. Validating vignette and conjoint survey experiments against real-world behavior. *Proc. Natl Acad. Sci. USA* **112**, 2395–2400 (2015).

34. Caplan, A. L. Stigma, vaccination, and moral accountability. *Lancet* **399**, 626–627 (2022).

35. Lyngse, F. P. et al. Household transmission of the SARS-CoV-2 Omicron variant in Denmark. *Nat. Commun.* **13**, 5573 (2021).

36. Hogan, A. M. et al. The value of vaccine booster doses to mitigate the global impact of the Omicron SARS-COV-2 variant. Preprint at *medRxiv* https://doi.org/10.1101/2022.01.17.22269222 (2022).

37. *Myths and Facts About COVID-19 Vaccines* (CDC, 2021); https://www.cdc.gov/coronavirus/2019-ncov/vaccines/facts.html.

38. Gelfand, M. J. et al. Differences between tight and loose cultures: a 33-nation study. *Science* **332**, 1100–1104 (2011).

39. Bollyky, T. J. et al. Pandemic preparedness and COVID-19: an exploratory analysis of infection and fatality rates, and contextual factors associated with preparedness in 177 countries, from Jan 1, 2020, to Sept 30, 2021. *Lancet* **399**, 1489–1512 (2022).

40. Gelfand, M. J. et al. The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. *Lancet Planet. Health* **5**, e135–e144 (2021).

41. Crandall, C. S., Eshleman, A. & O'brien, L. Social norms and the expression and suppression of prejudice: the struggle for internalization. *J. Pers. Soc. Psychol.* **82**, 359–378 (2002).

42. Pettigrew, T. F. Intergroup contact theory. *Ann. Rev. Psychol.* **49**, 65–85 (1998).

43. *United States Country Report* (Freedom House, 2022); https://freedomhouse.org/country/united-states/freedom-world/2022.

44. Aw, J., Seng, J. J. B., Seah, S. S. Y. & Low, L. L. COVID-vaccine hesitancy—a scoping review of literature in high-income countries. *Vaccines* **9**, 900 (2021).

45. Harrison, J., Berry, S., Mor, V. & Gifford, D. "Somebody like me": understanding COVID-19 vaccine hesitancy among staff in skilled nursing facilities. *J. Am. Med. Dir. Assoc.* **22**, 1133–1137 (2021).

46. *Statusreport 2021—COVID-19 Vaccination* (Sundhedsstyrelsen, 2021); https://www.sst.dk/-/media/Udgivelser/2021/Corona/Vaccination/Statusrapport/22_-statusraport-data-per-23_-aug.ashx.

47. Murphy, J. et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. *Nat. Commun.* **12**, 29 (2021).

48. Razai, M. S., Osama, T., McKechnie, D. G. & Majeed, A. *COVID-19 Vaccine Hesitancy Among Ethnic Minority Groups*. *BMJ* **372**, n513 (2021).

49. Sun, S., Lin, D. & Operario, D. Interest in COVID-19 vaccine trials participation among young adults in China: willingness, reasons for hesitancy, and demographic and psychosocial determinants. *Prev. Med. Rep.* **22**, 101350 (2021).

50. Robertson, E. et al. Predictors of COVID19 vaccine hesitancy in the UK household longitudinal study. *Brain Behav. Immun.* **94**, 41–50 (2021).

51. Kampf, G. COVID-19: stigmatising the unvaccinated is not justified. *Lancet* **398**, 1871 (2021).

52. Schmitt, M. T., Branscombe, N. R., Postmes, T. & Garcia, A. The consequences of perceived discrimination for psychological well-being: a meta-analytic review. *Psychol. Bull.* **140**, 921–948 (2014).

53. Branscombe, N. R., Schmitt, M. T. & Harvey, R. D. Perceiving pervasive discrimination among African Americans: implications for group identification and wellbeing. *J. Pers. Soc. Psychol.* **77**, 135–149 (1999).

54. Williamson, L. D., Smith, M. A. & Bigman, C. A. Does discrimination breed mistrust? Examining the role of mediated and non-mediated discrimination experiences in medical mistrust. *J. Health Commun.* **24**, 791–799 (2019).

55. Boon-Falleur, M., Grandin, A., Baumard, N. & Chevallier, C. Leveraging social cognition to promote effective climate change mitigation. *Nat. Clim. Change* **12**, 332–338 (2022).

56. Newman, N. et al. *Reuters Institute Digital News Report 2021* (Reuters, 2021).

57. *Survey Tool and Guidance: Rapid, Simple, Flexible Behavioural Insights on COVID-19: 29 July 2020* Technical Report (WHO, 2020).

58. Arechar, A. A. et al. Understanding and reducing online misinformation across 16 countries on six continents. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/a9frz (2022).

59. Awad, E. et al. The moral machine experiment. *Nature* **563**, 59–64 (2018).

60. Gachter, S. & Schulz, J. F. Intrinsic honesty and the prevalence of rule violations across societies. *Nature* **531**, 496–499 (2016).

61. Coppock, A., Leeper, T. J. & Mullinix, K. J. Generalizability of heterogeneous treatment effect estimates across samples. *Proc. Natl Acad. Sci. USA* **115**, 12441–12446 (2018).

62. Mullinix, K. J., Leeper, T. J., Druckman, J. N. & Freese, J. The generalizability of survey experiments. *J. Exp. Polit. Sci.* **2**, 109–138 (2015).

63. Hainmueller, J., Hopkins, D. J. & Yamamoto, T. Causal inference in conjoint analysis: understanding multidimensional choices via stated preference experiments. *Polit. Anal.* **22**, 1–30 (2014).

64. Blair, G., Imai, K. & Zhou, Y.-Y. Design and analysis of the randomized response technique. *J. Am. Stat. Assoc.* **110**, 1304–1319 (2015).

Acknowledgements

This research has been supported by the Carlsberg Foundation’s grant CF20-0044 and the Danish National Research Foundation’s grant DRNF144. We thank F. Hyldgaard for research assistance; A. Romano, J. Rasmussen, G. Simonovits, D. Hangartner, S. Lehmann, K. Lippert-Rasmussen, S. F. Midtgaard, V. M. L. Pedersen and M. Levendusky for comments; and L. Pradella, G. Spadaro, L. Sztankevics, A. Marie, C. M. Bosancianu, O. Ghajjou, M. C. Rizzato and B.-c. Zhang for foreign language assistance.

Author information

Authors and Affiliations

1. Department of Political Science, Aarhus University, Aarhus, Denmark

   Alexander Bor, Frederik Jørgensen & Michael Bang Petersen

2. Democracy Institute, Central European University, Budapest, Hungary

   Alexander Bor

3. Centre for the Experimental-Philosophical Study of Discrimination, Aarhus University, Aarhus, Denmark

   Michael Bang Petersen

Contributions

A.B., F.J. and M.B.P. designed the experiments. A.B. and M.B.P. conducted the experiments. A.B. analysed data with input from F.J. and M.B.P.; A.B. and M.B.P. wrote the paper. All of the authors approved the final manuscript.

Corresponding authors

Correspondence to Alexander Bor or Michael Bang Petersen.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Christian Crandall, John Roozenbeek and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Participants think about social interactions substantially more than about monetary transactions with vaccination outgroups.

Dots denote means, errorbars denote 95% confidence intervals. N = 1,448. See more details in Supplementary Information K.

Extended Data Fig. 2 Proportion indicating unhappiness if a vaccination outgroup member married into their family.

The plot contrasts a standard direct question (in red) to a question implemented with the forced response technique (in blue). Both methods indicate identical conclusions. Errorbars denote 95% confidence intervals. N = 1,448. See more details in Supplementary Information L.

Extended Data Fig. 3 The role of the three proximate outcomes: fear of infection, untrustworthiness, and incompetence.

Panel A displays average marginal component effects of target vaccination status on each of the three outcomes splitting on respondent vaccination status ($N_{vaccinated}$ = 54,054 and $N_{unvaccinated}$ = 10,386). Panel B displays the marginal regression coefficients from regressing prejudice on the three proximate variables simultaneously, while including respondent fixed effects. Errorbars denote 90% and 95% confidence intervals. See more details in Supplementary Information C.

Extended Data Fig. 4 Relationship between prejudice against unvaccinated individuals and policy stringency.

Estimates of exclusionary attitudes (based on average marginal component effects). Blue line denotes best fitting linear regression line, and grey lines denotes a loess curve. (Total N = 64,440 observations in 21 countries). Policy stringency is based on the Oxford COVID-19 Government Response Tracker by Hale et al (2020). See more details in Supplementary Information G.

Extended Data Fig. 5 Prejudice towards vaccination outgroups conditional on outgroup contact.

The left panel shows average marginal component effects and demonstrates that prejudice is highest among respondents with no contact at all, and smallest among those with most contacts. The right panel displays marginal means to offer more nuance. Errorbars denote 95% confidence intervals. Total N = 18,270. See details in Supplementary Information J.

Extended Data Fig. 6 Heterogeneities in exclusionary attitudes by education and income in each country.

Average marginal component effects (AMCEs) for exclusionary attitudes against unvaccinated individuals by country across lower and higher educated respondents (left panel), and poor and not poor respondents (right

panel). Estimates are based on Bayesian multilevel regression models. Error bars denote 90 and 95% credible intervals. Total N = 64,440. See more details in Supplementary Information D.

Extended Data Fig. 7 Simulations show that even if no offline citizen shows any exclusionary attitudes our main conclusions remain unchanged.

Original average marginal component effect estimates of exclusionary attitudes against unvaccinated individuals (orange dots) and simulations estimating the same under maximal bias from non-online populations (purple dots). Errorbars display 90% and 95% confidence intervals. Total N = 54,054. See more details in Supplementary Information O.2.

Extended Data Table 1 Study 1 – Sample demographics by country
Extended Data Table 2 Study 1 – Attributes and levels in the conjoint experimenta
Extended Data Table 3 Studies 2 & 3 – Attributes and levels in the conjoint experiment

Supplementary information

Supplementary Information

Supplementary A–O, including 48 Supplementary Figs. and 6 Supplementary Tables.

Reporting Summary

Peer Review File

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Bor, A., Jørgensen, F. & Petersen, M.B. Discriminatory attitudes against unvaccinated people during the pandemic. *Nature* **613**, 704–711 (2023). https://doi.org/10.1038/s41586-022-05607-y

- Received: 18 February 2022
- Accepted: 28 November 2022
- Published: 08 December 2022
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05607-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

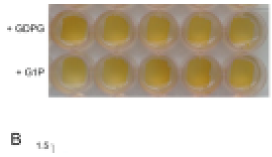
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

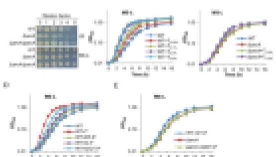
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05607-y



[A cAMP/CRP-controlled mechanism for the incorporation of extracellular ADP-glucose in Escherichia coli involving NupC and NupG nucleoside transporters]

19 October 2018

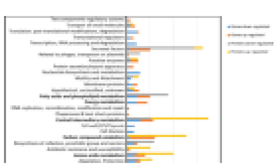
Goizeder Almagro, Alejandro M. Viale, … Javier Pozueta-Romero



[Transcriptional regulator ArcA mediates expression of oligopeptide transport systems both directly and indirectly in Shewanella oneidensis]

25 September 2019

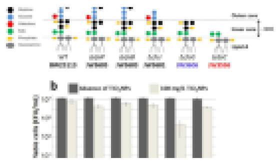
Huihui Liang, Yinting Mao, … Haichun Gao



[The absence of SigX results in impaired carbon metabolism and membrane fluidity in Pseudomonas aeruginosa]

21 November 2018

Maud Fléchard, Rachel Duchesne, … Sylvie Chevalier



Osmotic stress and vesiculation as key mechanisms controlling bacterial sensitivity and resistance to TiO2 nanoparticles

03 June 2021

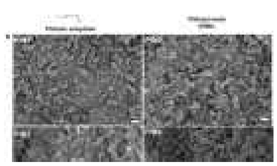
Christophe Pagnout, Angelina Razafitianamaharavo, … Jérôme F. L. Duval



Bacterial microcompartments

05 March 2018

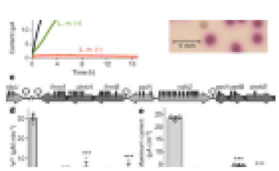
Cheryl A. Kerfeld, Clement Aussignargues, … Markus Sutter



A new way of producing pediocin in Pediococcus acidilactici through intracellular stimulation by internalized inulin nanoparticles

12 April 2018

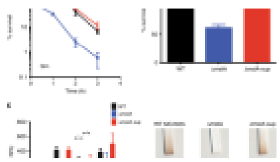
Whee-Soo Kim, Jun-Yeong Lee, … Chong-Su Cho



A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria

12 September 2018

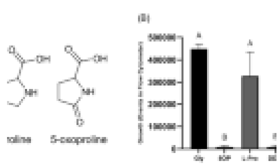
Samuel H. Light, Lin Su, … Daniel A. Portnoy



Transcription factor YcjW controls the emergency H2S production in E. coli

28 June 2019

Lyly Luhachack, Aviram Rasouly, … Evgeny Nudler



Characterizing 5-oxoproline sensing pathways of Salmonella enterica serovar typhimurium

24 September 2022

Einav Stern, Naama Shterzer & Erez Mills

- Article
- Open Access
- Published: 30 November 2022

Undecaprenyl phosphate translocases confer conditional microbial fitness

- Brandon Sit,
- Veerasak Srisuknimit,
- Emilio Bueno,
- Franz G. Zingl,
- Karthik Hullahalli,
- Felipe Cava &
- …
- Matthew K. Waldor

Show authors

Nature **volume 613**, pages 721–728 (2023)

- 4797 Accesses
- 3 Citations
- 81 Altmetric
- Metrics details

Abstract

The microbial cell wall is essential for maintenance of cell shape and resistance to external stressors[1]. The primary structural component of the cell wall is peptidoglycan, a glycopolymer with peptide crosslinks located outside of the cell membrane[1]. Peptidoglycan biosynthesis and structure are responsive to shifting environmental conditions such as pH and salinity[2,3,4,5,6], but the mechanisms underlying such adaptations are incompletely understood. Precursors of peptidoglycan and other cell surface

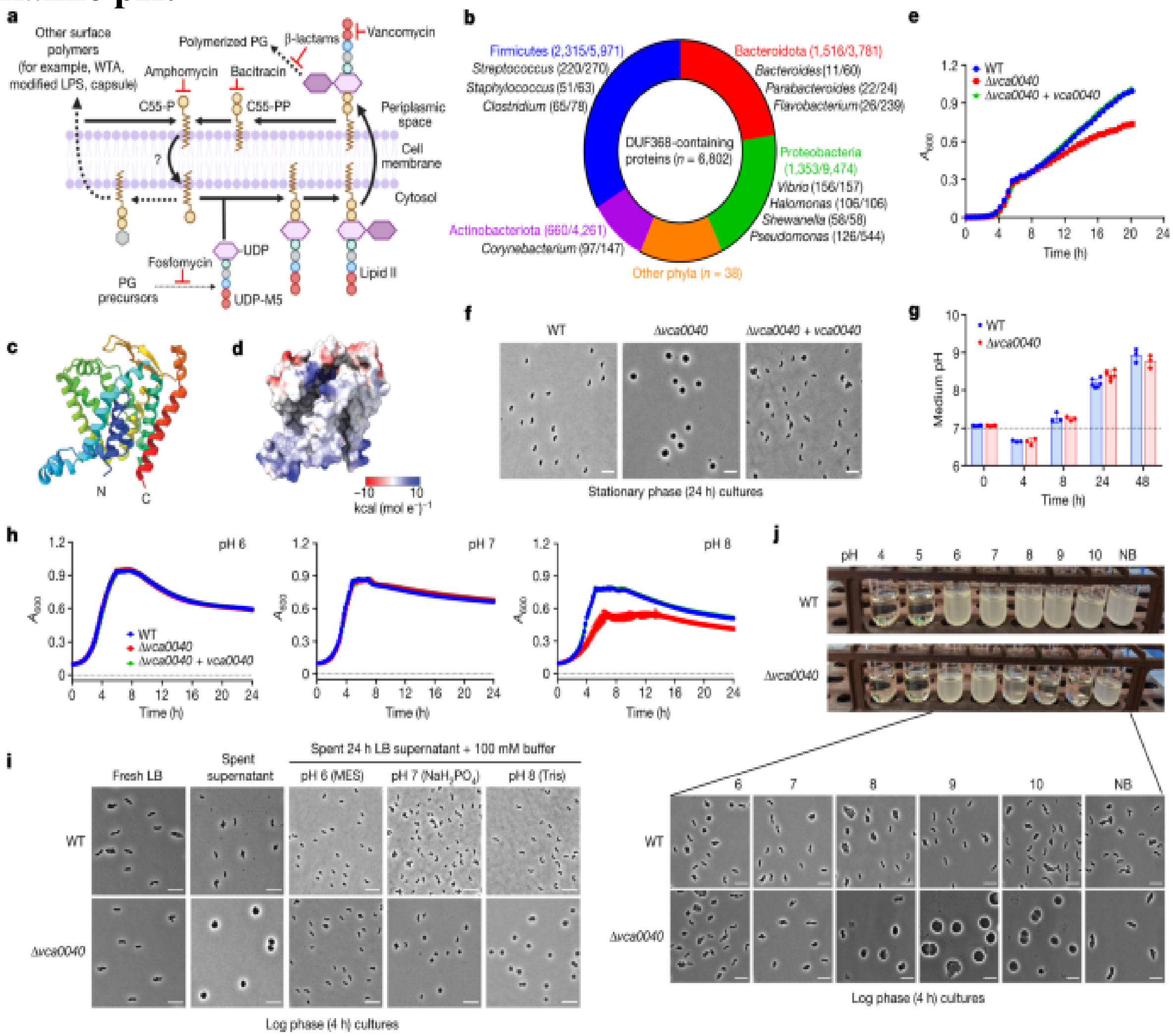
glycopolymers are synthesized in the cytoplasm and then delivered across the cell membrane bound to the recyclable lipid carrier undecaprenyl phosphate[7] (C55-P, also known as UndP). Here we identify the DUF368-containing and DedA transmembrane protein families as candidate C55-P translocases, filling a critical gap in knowledge of the proteins required for the biogenesis of microbial cell surface polymers. Gram-negative and Gram-positive bacteria lacking their cognate DUF368-containing protein exhibited alkaline-dependent cell wall and viability defects, along with increased cell surface C55-P levels. pH-dependent synthetic genetic interactions between DUF368-containing proteins and DedA family members suggest that C55-P transporter usage is dynamic and modulated by environmental inputs. C55-P transporter activity was required by the cholera pathogen for growth and cell shape maintenance in the intestine. We propose that conditional transporter reliance provides resilience in lipid carrier recycling, bolstering microbial fitness both inside and outside the host.

Main

Phosphorylated undecaprenyl (C55) lipids have an essential role as recyclable carrier molecules in microbial cell surface glycopolymer biogenesis[7]. During peptidoglycan biosynthesis, the sugar-linked pentapeptide subunits of peptidoglycan that are assembled in the bacterial cytoplasm are covalently linked to C55-P for transmembrane transport[7] (Fig. 1a). After the carrier-linked peptidoglycan precursor (lipid II) is moved from the inner to outer leaflet of the membrane by MurJ[8], subsequent incorporation of the muropeptide into polymerized peptidoglycan leaves behind C55 pyrophosphate (C55-PP). Carrier recycling is initiated by hydrolysis of C55-PP to C55-P by membrane-associated phosphatases including UppP (also known as BacA) and the PAP2-domain proteins PgpB, YbjG and LpxT[7]. Then, C55-P is presumably flipped back into the cytosolic face of the membrane to complete recycling. Preliminary structural studies have proposed that UppP may also function as a C55-P translocase[9,10], but the protein(s) responsible for C55-P internalization have not yet been identified. C55-P recycling is a key step in the biosynthesis of peptidoglycan as well as other cell surface glycopolymers, including wall teichoic acids, certain lipopolysaccharide modifications and capsules[7]. Given its wide-

ranging and critical role in cell surface maintenance, C55-P recycling is considered to be an important target for antimicrobial therapies, and naturally occurring antibiotics that inhibit this process have been identified[11].

Fig. 1: VCA0040 is required for cell shape maintenance of *V. cholerae* at alkaline pH.



a, Usage and recycling of C55-P in bacteria. Sites of antibiotic action relevant to Fig. 3 are indicated with red bars. Dashed arrows represent multiple enzymatic and/or transport steps. The grey hexagon represents variable sugars linked to C55-P for downstream glycopolymer assembly. LPS, lipopolysaccharide; PG, peptidoglycan. **b**, Conservation of DUF368 in selected bacterial phyla (upright) and genera (italic). The coloured segments and associated labels denote selected phyla with substantial DUF368

conservation. The fraction of sequenced unique clade genomes encoding a DUF368-containing protein is indicated. **c**,**d**, Predicted ribbon (**c**) and electrostatic surface coloured (**d**) structures of VCA0040. Colour scale, −10 to 10 kcal (mol e^{-}). **e**,**f**, Growth (**e**) and morphology (**f**) of wild-type (WT), Δ*vca0040* and Δ*vca0040* + *vca0040* (chromosomally complemented) *V. cholerae* in LB medium. **g**, Medium pH of *V. cholerae* overnight LB cultures. **h**, *V. cholerae* growth in M9 medium buffered to the indicated pH. **i**, The effects of buffered spent supernatant (cell-free medium after 24 h of growth at 30 °C from a 1:1,000 culture) on *V. cholerae* morphology in log phase. **j**, The effects of pH on *V. cholerae* growth (top) and morphology (bottom) during log phase in LB medium buffered with 50 mM Na_2HPO_4. NB, unbuffered medium. **e**,**g**,**h**, Data are mean ± s.d. from $n = 3$ cultures per strain or condition. **f**,**i**,**j**, Representative images from $n = 3$ cultures per strain or condition. Scale bars, 3 μm (**f**) and 5 μm (**i**,**j**).

Vibrio cholerae, the Gram-negative causal agent of the diarrheal disease cholera, is exposed to dramatic changes in pH and ion concentrations as it enters, transits and exits the host gastrointestinal tract. A variety of sensing and signalling networks enable *V. cholerae* to adapt to changing environments. For example, the pathogen responds to altered salinity and pH by using a sodium motive force (SMF) instead of a proton motive force (PMF) to power protein secretion and flagellar-dependent motility, and regulate virulence gene expression[12,13]. The peptidoglycan composition of *V. cholerae* is thought to be influenced by environmental inputs, but the effects of pH on *V. cholerae* cell wall biology are unknown[14].

DUF368 impacts *V. cholerae* alkaline fitness

A recent in vivo transposon-insertion sequencing screen for determinants of intestinal colonization in a contemporary *V. cholerae* clinical isolate identified numerous genes not previously linked to *V. cholerae* pathogenesis, including several loci of unknown function[15] (Extended Data Fig. 1a). One of these genes, *vca0040* (N900_RS14215), was selected for further study because similar datasets[16,17,18] suggested a universal requirement for this locus in *V. cholerae* intestinal colonization. VCA0040 is predicted to be a multi-pass inner membrane protein and contains the widely conserved

domain of unknown function DUF368 (also known as PF04018) (Fig. 1b and Extended Data Fig. 1b–e). The predicted structural model of VCA0040 has a large putative cleft, characteristic of domains with ligand binding and/or transport activity (Fig. 1c,d and Extended Data Fig. 2). A *V. cholerae* strain lacking VCA0040 (Δ*vca0040*) exhibited a stationary phase growth defect (Fig. 1e and Extended Data Fig. 3a) that was accompanied by a distinct cell shape phenotype, where Δ*vca0040* cells became large spheres (Fig. 1f and Extended Data Fig. 3b,c). These spherical *V. cholerae* were viable and gave rise to normal rod-shaped daughter cells (Extended Data Fig. 3d).

We reasoned that a stationary phase-specific factor could be triggering the shape defect in Δ*vca0040* cells. Accordingly, exposure of exponentially growing Δ*vca0040* cells to cell-free spent supernatants from stationary phase cultures rapidly induced sphere formation (Extended Data Fig. 3e). Heat-labile and high-molecular-weight factors, as well as d-amino acids—which modulate stationary phase peptidoglycan composition[19]—were excluded as candidates (Extended Data Fig. 3e,f). In LB cultures, *V. cholerae* entry into stationary phase is accompanied by media alkalinization (Fig. 1g). Since minimal shape defects were observed in neutral buffered M9 medium (Extended Data Fig. 3b), we hypothesized that Δ*vca0040* cells were alkaline-sensitive. Buffering M9 media to different pH values revealed a growth defect at pH 8, but not at pH 6 or pH 7, and acidification of the cell-free spent supernatant ablated the sphere-induction phenotype (Fig. 1h,i). In buffered LB, Δ*vca0040* cells exhibited growth and cell shape defects at pH above 8, but not at pH values of up to 7 (Fig. 1j). Thus, the DUF368-containing protein VCA0040 is required for *V. cholerae* cell shape integrity and growth in alkaline conditions.

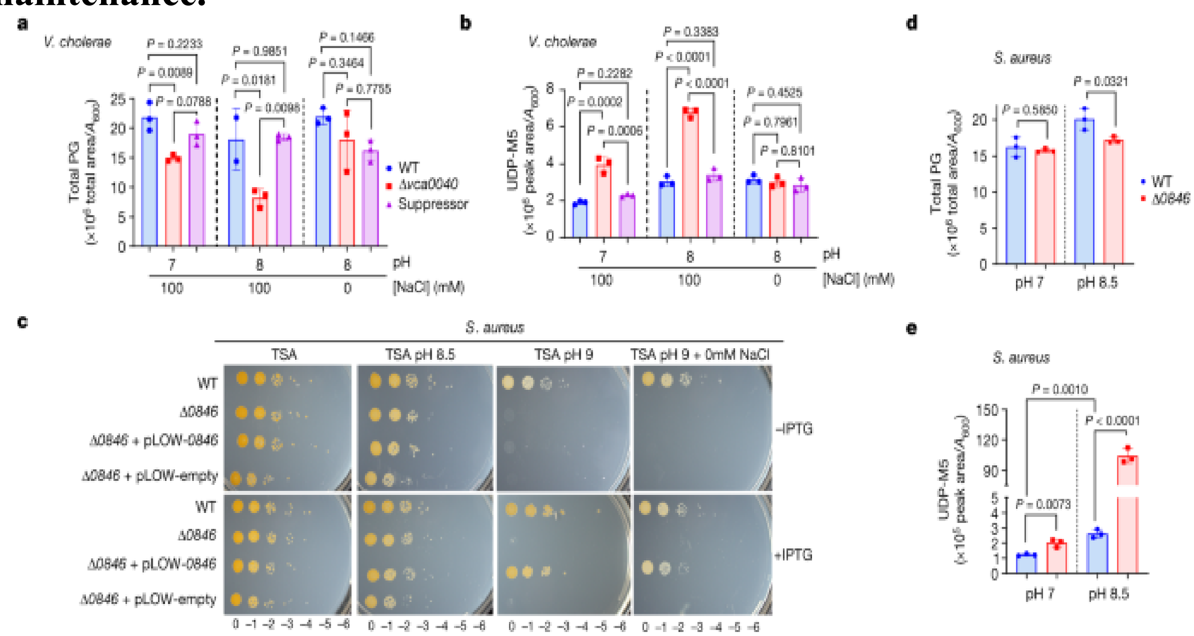
Altered peptidoglycans in DUF368 mutants

DUF368 domains are almost universally conserved across the Vibrionaceae and are present in thousands of additional Gram-negative, Gram-positive and archaeal species that inhabit a wide range of niches (Supplementary Tables 1 and 2). Heterologous expression of Gram-positive (*Staphylococcus aureus*) or archaeal (*Haloferax volcanii*) DUF368 homologues at least partially complemented the alkaline growth defect of Δ*vca0040 V. cholerae*

cells (Extended Data Fig. 4a,b). Both homologues showed strong predicted structural similarity to VCA0040, suggesting that DUF368-containing protein function is conserved not only between Gram-negative and Gram-positive species, but across microbial kingdoms (Extended Data Fig. 2).

Since peptidoglycan is required for maintenance of bacterial cell shape, we hypothesized that DUF368 functions affect the cell wall. The Δ*vca0040* mutant had 1.5–2× less peptidoglycan than the wild type, and modest cross-linking defects with concurrent accumulation of the peptidoglycan precursor UDP-*N*-acetylmuramyl pentapeptide (UDP-M5), suggestive of a peptidoglycan biosynthesis defect upstream of cross-linking (Fig. 2a,b and Extended Data Fig. 4c–e). These phenotypes were present at neutral pH and were exacerbated by exposure to alkaline conditions. Consistent with the idea that DUF368 function is conserved, *S. aureus* lacking *SAOUHSC_00846* also had an alkaline growth defect, reduced peptidoglycan quantity, accumulation of UDP-M5 and decreased cross-linking (Fig. 2c–e and Extended Data Fig. 4f,g). We concluded from these data that DUF368-containing proteins are conditionally required to maintain peptidoglycan production and composition.

Fig. 2: DUF368 function is conserved and necessary for peptidoglycan maintenance.



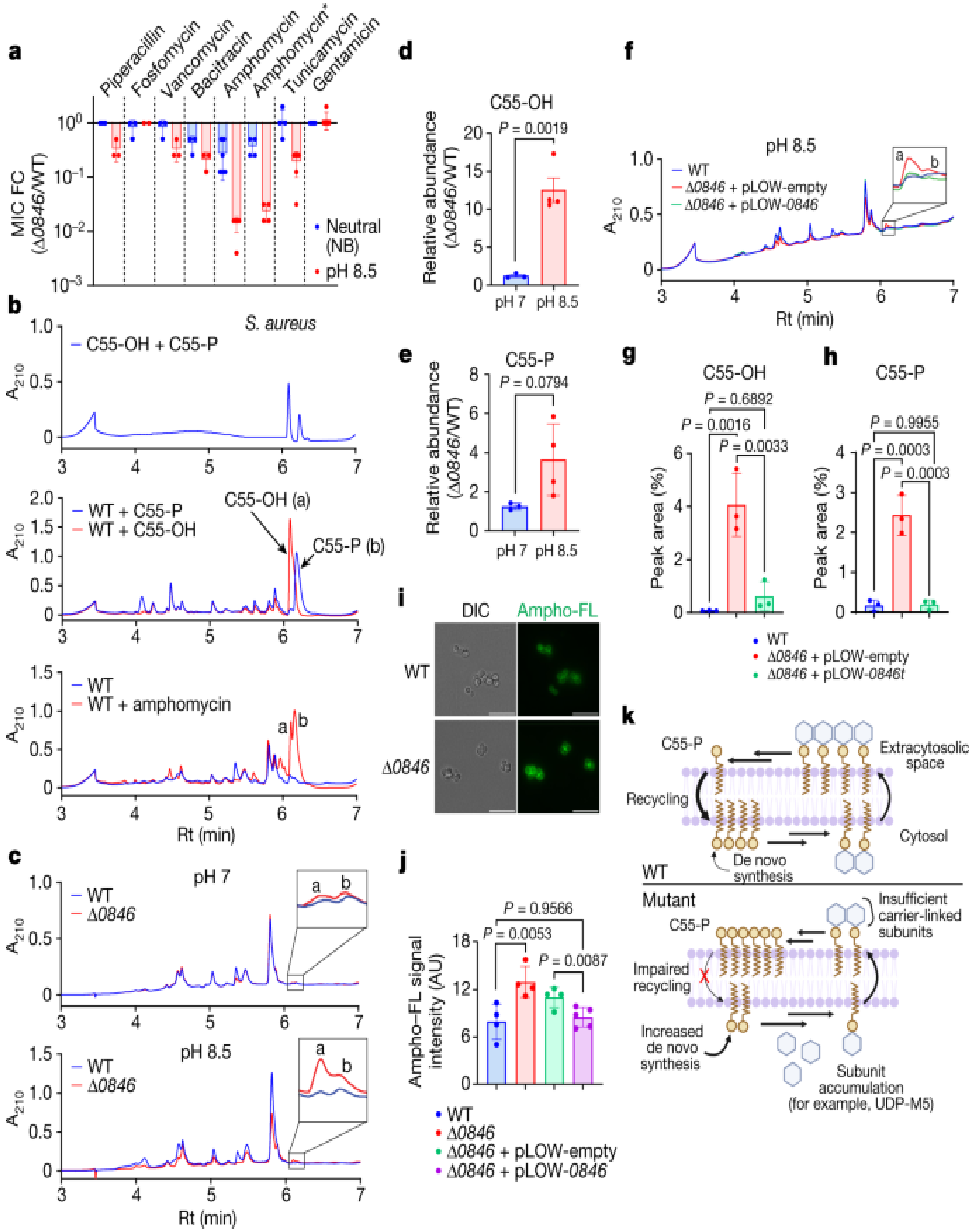
a,**b**, Total peptidoglycan (**a**) and intracellular UDP-M5 (**b**) in *V. cholerae* grown in M63 minimal medium at the indicated pH. Suppressor, Δ*vca0040*/*ΔsecDF1*. **c**, Alkaline growth of Δ*SAOUHSC_00846* (Δ*0846*) *S. aureus* in the indicated conditions and with the indicated expression vectors. Representative results from $n = 3$ independent experiments per condition. **d**,**e**, Total peptidoglycan (**d**) and intracellular UDP-M5 (**e**) in *S. aureus* grown in tryptic soy broth (TSB) with 100 mM bicine at the indicated pH. **a**,**b**,**d**,**e**, $n = 3$ cultures per strain or condition; data are mean ± s.d. **a**,**b**, One-way ANOVA with Tukey's multiple comparison test. **d**,**e**, Unpaired Student's two-tailed *t*-test.

Impaired C55-P recycling in DUF368 mutants

Two additional observations focused our investigation of DUF368 domains on cell wall assembly. First, we observed that although most DUF368-containing proteins consist primarily of one or two DUF368 domains, several microorganisms encode dual-domain proteins with DUF368–PAP2, PAP2–DUF368 or DUF368–BacA architectures (Supplementary Table 3). Second, during sphere formation in Δ*vca0040 V. cholerae*, there was an approximately tenfold induction of *pgpB*, which encodes a C55-PP phosphatase[20] (Extended Data Fig. 5a–d and Supplementary Table 4). Despite these associations of DUF368 with C55-related processes, the Δ*vca0040* mutant exhibited only minor (up to 4×) increases in minimum inhibitory concentrations (MICs) of an array of peptidoglycan-targeting antibiotics, probably because most cell wall-acting compounds cannot penetrate the Gram-negative outer membrane (Supplementary Table 5). Accordingly, MIC screening in *S. aureus* revealed that the Δ*SAOUHSC_00846* mutant was far more sensitive (more than 64×) to amphomycin than the wild type in alkaline conditions (Fig. 3a and Supplementary Table 5). Amphomycin is a Ca^{2+}-dependent lipopeptide antibiotic that specifically binds C55-P in the outer leaflet of the cytoplasmic membrane and inhibits its recycling[21,22,23] (Fig. 1a). Consistent with our previous data (Fig. 2d,e), amphomycin is known to induce UDP-M5 accumulation and decrease peptidoglycan cross-linking in *S. aureus*[24]. The Δ*SAOUHSC_00846* mutant was additionally moderately sensitized to tunicamycin, which inhibits the first committed and C55-P-dependent steps of wall teichoic acid and peptidoglycan synthesis as well as bacitracin,

which binds C55-PP and prevents C55-P generation on the extracytosolic face of the membrane[25,26] (Figs. 1a and 3a). MICs for other peptidoglycan-targeting molecules such as fosfomycin were minimally changed (Fig. 3a and Supplementary Table 5). To test DUF368 function in a Gram-negative organism, we heterologously expressed *SAOUHSC_00846* or *vca0040* in the outer membrane-permeable—and thus amphomycin-sensitized—*Escherichia coli* strain *lptD4213*[27] (Supplementary Table 5). Expression of either DUF368-containing protein in *lptd4213 E. coli* conferred amphomycin resistance (Supplementary Table 5). Since *E. coli* lacks an endogenous DUF368-containing protein, this result indicates that DUF368 function is sufficient for resistance to a surface C55-P-targeting antibiotic.

Fig. 3: C55-P recycling is impaired in bacteria lacking DUF368-containing proteins.



a, Fold change (FC) in MIC for Δ*SAOUHSC_00846 S. aureus* in the indicated conditions. Amphomycin*, amphomycin with 1 mM $CaCl_2$. **b**, Ultra high performance liquid chromatography (UPLC) traces of purified C55-OH (a) and C55-P (b) standards (top), wild-type *S. aureus* lipid extracts

spiked with each standard (middle) or wild-type *S. aureus* treated with 50 μg ml^{-1} amphomycin (bottom). Rt, retention time. **c**, UPLC traces of lipid extracts of wild-type or mutant *S. aureus* grown in TSB pH 7 (top) or pH 8.5 (bottom). **d**,**e**, Relative abundance of C55-OH (**d**) and C55-P (**e**) under same conditions as in **c**. **f**, UPLC trace of lipid extracts of complemented Δ*SAOUHSC_00846 S. aureus* grown in TSB pH 8.5. **g**,**h**, Raw peak proportions of C55-OH (**g**) and C55-P (**h**) in each strain grown at pH 8.5. **i**, *S. aureus* stained with ampho-FL at pH 8.5. Scale bars, 5 μm. DIC, differential interference contrast image. **j**, Quantification of ampho-FL signal. Symbols represent the mean ampho-FL intensity of 1 to 8 individual bacterial clusters in independent cultures. **k**, Proposed model of disrupted C55-P recycling and the associated consequences in bacteria lacking C55-P translocase activity. **b**,**c**,**f**,**i**, Representative data from $n = 3$ cultures per strain or condition. **d**,**e**,**g**,**h**,**j**, Data are mean ± s.d. from $n = 3$–5 cultures per strain or condition. Each point is an independent culture. **d**,**e**, Unpaired student's two-tailed *t*-test. **g**,**h**,**j**, One-way ANOVA with Tukey's multiple comparison test.

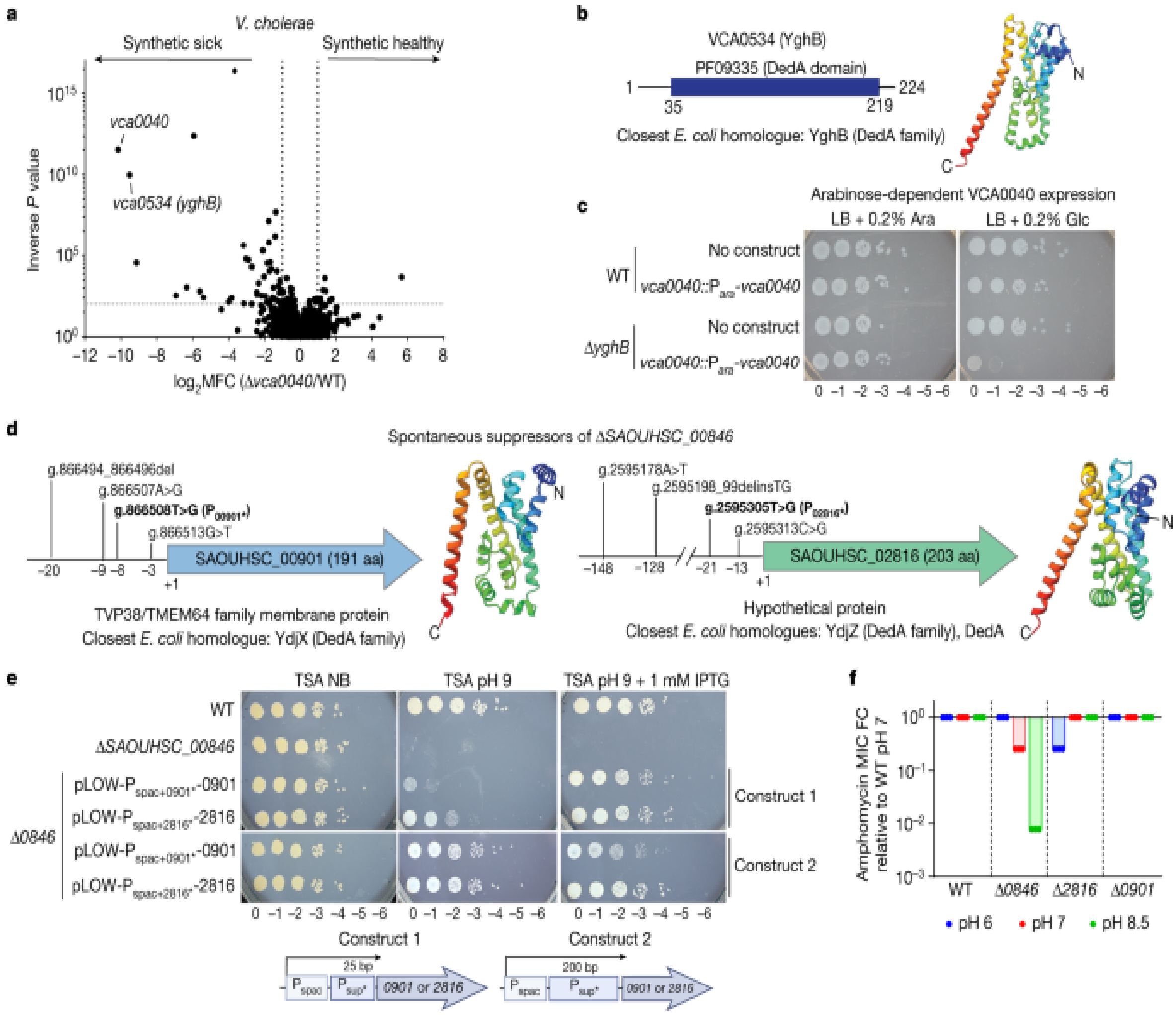
The peptidoglycan composition and antibiotic susceptibility data indirectly suggested that C55-P recycling was impaired in the Δ*SAOUHSC_00846* mutant, specifically in C55-P re-internalization. To further investigate this idea, we quantified C55 species in membrane lipid extracts from wild-type and Δ*SAOUHSC_00846* cultures. Amphomycin treatment of wild-type *S. aureus* led to marked accumulation of both C55-P and the Gram-positive-restricted C55-P precursor C55-OH[28] (undecaprenol, also known as UndOH) (Fig. 3b). This phenotype was replicated in Δ*SAOUSHC_00846 S. aureus* cells grown in alkaline media, but not in neutral conditions (Fig. 3c–e) or in a complemented strain (Fig. 3f–h), providing direct evidence that C55-P homeostasis is linked to *SAOUSHC_00846*. We also observed similar alkaline-dependent increases in C55-P (but not C55-OH) in Δ*vca0040 V. cholerae* (Extended Data Fig. 5e,f). To specifically quantify C55-P in the outer leaflet of the cell membrane, we synthesized fluorescein-conjugated amphomycin (ampho-FL) and used it to label live bacterial cells[29]. Ampho-FL-labelled Δ*SAOUHSC_00846 S. aureus* exhibited increased signal relative to wild-type bacteria in alkaline media, demonstrating that surface C55-P levels are conditionally elevated in the mutant (Fig. 3i,j and Extended Data Fig. 5g,h). Together, these observations suggest that the alkaline

accumulation of surface C55-P in DUF368-containing protein mutants provokes a compensatory—but ultimately insufficient—increase in C55-P synthesis, resulting in surface glycopolymer production defects and impaired growth (Fig. 3k). These findings are consistent with a model in which DUF368-containing proteins are C55-P recycling transporters that are active in alkaline conditions.

Genetic interactions between DUF368 and DedA

As C55-P recycling is thought to be an essential function and DUF368 mutants were conditionally viable, we next carried out a synthetic transposon screen to define the genetic network of *vca0040*. Many of the identified synthetic sick or lethal interactions were related to cell envelope homeostasis (Fig. 4a and Supplementary Table 6). The strongest interaction of *vca0040* was with *N900_RS16280* (also known as *vca0534*), which encodes a homologue of the *E. coli* DedA family member YghB (Fig. 4b). *V. cholerae* express two other DedA proteins, but these were not hits in this screen. DedA family transmembrane proteins (also known as PF09335 (SNARE-associated Golgi protein)) are conserved in all three domains of life, are generally required for cell envelope homeostasis, and are suspected to mediate PMF-dependent transport, but their specific substrates are unknown[30,31]. Similar to DUF368, although most DedA-containing sequences encode only that domain, there are many instances of domain architectures in which DedA is fused to the C55-PP phosphatase domain PAP2 (Supplementary Table 3). The synthetic lethality of *vca0040* and *yghB* was validated by genetic depletion and reciprocal synthetic transposon screening (Fig. 4c, Extended Data Fig. 6a and Supplementary Table 7), and overexpression of YghB rescued the alkaline defect of Δ*vca0040 V. cholerae* (Extended Data Fig. 6b).

Fig. 4: DedA family members are additional C55-P translocase candidates.



a, A synthetic transposon screen in Δ*vca0040 V. cholerae*, showing $\log_2$ mean read fold changes (MFC; threshold ±2) and inverse Mann–Whitney *U* test *P*-values (threshold 100). Pooled data from two independent transposon libraries. **b**, Domain content and predicted structure of *V. cholerae* VCA0534 (also known as YghB). **c**, Growth of Δ*yghB V. cholerae vca0040* depletion strains. Ara, arabinose; Glc, glucose. **d**, Spontaneous suppressors of Δ*SAOUHSC_00846* map to two *S. aureus* DedA family proteins, with predicted structures (right) and specific promoter mutations selected for validation (bold). **e**, Rescue of Δ*SAOUHSC_00846* by expression of *SAOUHSC_00901* (*0901*) or *SAOUHSC_2816* (*2816*) from hybrid promoters containing the IPTG-inducible promoter P_{spac} linked to 25-bp (construct 1) or 200-bp (construct 2) native promoter stretches with the indicated suppressor mutations (*). **f**, Fold change in amphomycin MIC for wild-type and mutant *S. aureus* relative to the MIC for the wild type at pH 7

(150 μg ml^{-1}). **c**,**e**, Representative results from $n = 3$ independent experiments per condition. **f**, Data are mean ± s.d. from $n = 3$ cultures per strain or condition.

Interactions between DUF368 and DedA proteins were also observed in *S. aureus*. In alkaline-enriched spontaneous suppressors of Δ*SAOUHSC_00846 S. aureus*, we found frequent promoter (and no coding region) mutations in the genes encoding the two *S. aureus* DedA proteins (SAOUHSC_00901 and SAOUHSC_02816) (Fig. 4d). Incorporating these mutations into an isopropyl β-d-1-thiogalactopyranoside (IPTG)-inducible system expressing either DedA protein rescued Δ*SAOUHSC_00846* even without induction, suggesting that the isolated mutations increase gene expression and compensate for the absence of *SAOUHSC_00846*, as in *V. cholerae* (Fig. 4e). These genetic studies suggest that DedA family members are also required for C55-P translocation, in agreement with recent reports that eukaryotic DedA proteins are associated with lipid transport and bacterial DedA mutants are defective in C55 carrier-dependent lipopolysaccharide modifications[31,32,33].

We hypothesized that similar to DUF368-containing proteins, DedA family members function conditionally. Although deletion of *SAOUHSC_00901* did not affect amphomycin MICs, the Δ*SAOUHSC_02816* mutant exhibited moderate acid-dependent amphomycin sensitivity, strengthening the association of DedA family members with C55-P translocation and suggesting that they function at lower pH ranges than their DUF368-containing counterparts (Fig. 4f). The sensitivity of Δ*SAOUHSC_02816* to other antibiotics was unaltered at any pH, highlighting the specificity of conditional amphomycin sensitization (Supplementary Table 5). The concept of pH-dependent translocase contributions is also supported by the alkaline amphomycin sensitivity of *lptd4213 E. coli*, which has only DedA paralogues and no cognate DUF368-containing protein (Supplementary Table 5). A strain lacking *yghB* and *vca0040* was viable at acidic pH, suggesting at least one other *V. cholerae* protein can carry out C55-P translocation, or that C55-P may spontaneously traverse the membrane in acidic conditions (Extended Data Fig. 6c).

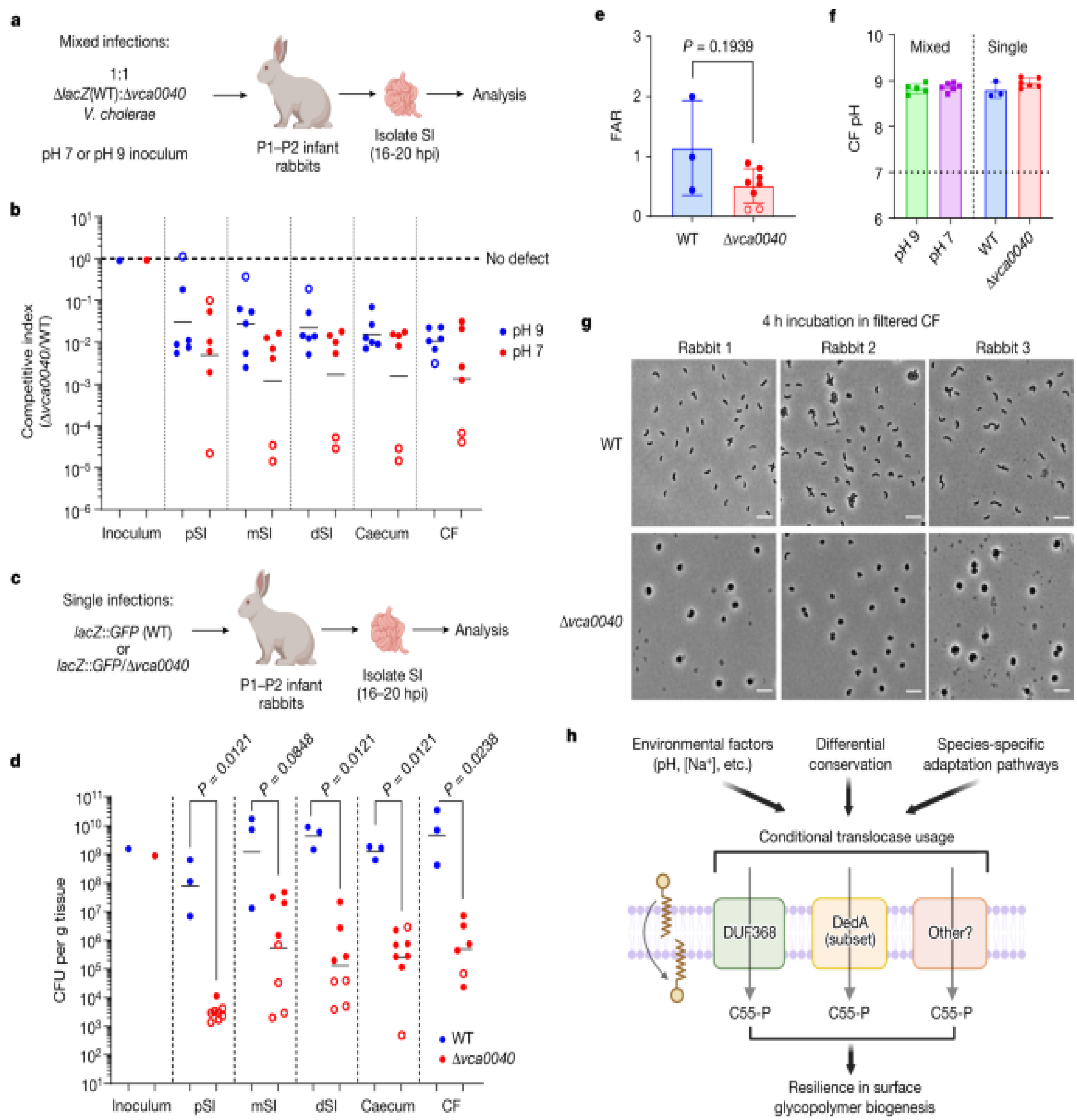
Na^+ shapes C55-P translocase biology

Analyses of spontaneous suppressors of Δ*vca0040 V. cholerae*, which have distinctive colony morphology (Extended Data Fig. 7a) revealed that the requirement of VCA0040 for *V. cholerae* fitness is controlled by Na^+ concentration in addition to pH. Whole-genome sequencing of suppressor colonies lacking cell shape defects revealed four suppressor genes, three of which (*secD1, secF1* and *ppiD*) function in Sec-mediated protein secretion (Extended Data Fig. 7b). The fourth—*yfgO*—encodes a protein of unknown function from the AI-2E transporter family, members of which have recently been implicated in peptidoglycan biosynthesis and Na^+/H^+ antiport[34,35]. Deletions of suppressor genes rescued the shape and peptidoglycan defects of Δ*vca0040 V. cholerae*, even though stationary culture pH was alkaline in all suppressors (Fig. 2a,b and Extended Data Fig. 7c,d). In *V. cholerae*, SecD–SecF is duplicated as SecD1–SecF1 and SecD2–SecF2, ancillary complexes that differentially couple SecYEG activity to the SMF or PMF, respectively[36] (Extended Data Fig. 7e). The activity of at least one SecD–SecF pair was required for *V. cholerae* viability, and deletion of *secD2* and *secF2* did not ameliorate the stationary phase cell shape defect of Δ*vca0040* (Extended Data Fig. 7c,f,g). We initially hypothesized that altered Sec substrate specificity could explain the identification of SecD1 and SecF1 as suppressors. However, the Δ*secDF1 V. cholerae* proteome was unchanged from that of wild-type cells, aside from the expected increases in SecD2 and SecF2, suggesting that the suppressive effects of *secD1* and *secF1* loss were attributable to reduced Na^+ flux (Extended Data Fig. 7h and Supplementary Table 8). Indeed, increasing Na^+ but not K^+ concentration induced alkaline growth and peptidoglycan defects in Δ*vca0040 V. cholerae* (Fig. 2a,b and Extended Data Fig. 7i–k). The requirement of increased Na^+ concentration for the alkaline defect of DUF368-deficient *V. cholerae* appeared to be species-specific, as Δ*SAOUHSC_00846 S. aureus*, although alkaline-sensitive, was not similarly rescued by Na^+ depletion (Fig. 2c). Of note, unlike Δ*vca0040*, a Δ*yghB V. cholerae* mutant was not able to grow without Na^+ at pH values above 6 (Extended Data Fig. 7l,m), suggesting that the presence of Na^+ divergently affects distinct C55-P translocases. Together, these data illustrate that multiple environmental inputs can control C55-P translocase requirements and potentially function in different microorganisms.

DUF368 enables *V. cholerae* pathogenesis

We used an infant (postnatal day (P)1–P2) rabbit model that mimics severe human cholera[37] to investigate the role of VCA0040 in pathogenesis. Typically, this model uses alkaline inocula to buffer the acidic stomach environment. To exclude inoculum effects, we carried out mixed infections with inocula comprising a 1:1 mix of wild-type and Δ*vca0040 V. cholerae* at pH 7 or pH 9 (Fig. 5a). Regardless of inoculum pH, the *Δvca0040* mutant had a profound competitive defect (100–1,000x) compared with the wild type (Fig. 5b). The pH of the inoculum did not affect total colonization or accumulation of diarrhea-like caecal fluid, the primary marker of disease in this model (Extended Data Fig. 8a,b). In single infections with fluorescent wild-type or Δ*vca0040 V. cholerae*, the mutant exhibited severe colonization defects (around 1,000-fold) in all intestinal segments (Fig. 5c,d). These values overestimate the true colonization of the mutant, because rabbits from each group were co-housed to exclude litter effects, enabling substantial transmission from wild-type-infected to Δ*vca0040*-infected rabbits (Extended Data Fig. 8c). Transmission probably explains why caecal fluid accumulation in Δ*vca0040*-infected rabbits was lower than in wild-type rabbits but did not reach statistical significance (Fig. 5e). Caecal fluid collected from both mixed and singly infected animals was strongly alkaline (pH 8.5–9) (Fig. 5f). Imaging of caecal fluid samples from Δ*vca0040*-infected rabbits revealed large, fluorescent, spherical cells similar to those observed in alkaline cultures of the Δ*vca0040* mutant (Extended Data Fig. 8d). Similarly, when freshly grown rod-shaped wild-type or Δ*vca0040 V. cholerae* were incubated in cell-free caecal fluid samples. Δ*vca0040*—but not wild-type—cells became spherical when exposed to caecal fluid, suggesting that VCA0040 function is needed to maintain *V. cholerae* shape in the alkaline in vivo milieu (Fig. 5g). An infection defect was not detected for Δ*SAOUHSC_00846 S. aureus* in a mouse intravenous infection–organ abscess model, in which the pathogen is unlikely to encounter alkaline environments (Extended Data Fig. 9).

Fig. 5: DUF368 is required for *V. cholerae* pathogenesis.



a,**b**, Schematic (**a**) and intestinal competitive indices (**b**) in mixed-infection models ($n = 6$ rabbits in each group). SI, small intestine. **c**–**e**, Schematic (**c**), number of intestinal colony-forming units (CFU) (**d**) and fluid accumulation ratio (FAR) (**e**) in single infections ($n = 3$ rabbits (wild type) and $n = 8$ rabbits (Δ*vca0040*)). CF, caecal fluid; pSI, proximal small intestine; mSI, medial small intestine; dSI, distal small intestine. Open circles indicate rabbits with limit of detection (LOD) measurements where the true CFU burden is at least (for upper LOD circles) or at most (for lower LOD circles) the plotted value. Note two Δ*vca0040* animals had insufficient caecal fluid accumulation for FAR calculation and were assigned LOD

values (arbitrary 25 µl volume). **b**,**d**, Data are geometric mean. **d**, Two-tailed Mann–Whitney *U* tests. **e**, Data are mean ± s.d. analysed with an unpaired Student's two-tailed *t*-test. **f**, pH of caecal fluid from infected animals. Data are mean ± s.d. **g**, Incubation of laboratory-grown *V. cholerae* with filtered caecal fluid samples from three different rabbits. Representative images from $n = 3$ ex vivo incubations with independent caecal fluid samples. Scale bars, 3 µm. **h**, Proposed model for microbial C55-P translocation and its integrated control by environmental inputs, the presence of other translocases, and non-translocase-related environmental adaptation mechanisms.

Source data

Discussion

We propose that DUF368-containing and DedA family proteins are C55-P translocases that provide functional resiliency in this crucial step of envelope maintenance (Fig. 5h). The identification of C55-P transport proteins provisionally fills a major gap in the C55 lipid carrier recycling pathway. As our genetic and phenotypic data do not definitively exclude transport-independent activities, biochemical and structural studies are required to confirm whether these two families indeed carry out this function. The apparent pH and ion dependency of the contributions of these candidate translocases to fitness suggests that these proteins could be energetically controlled, since a major effect of pH on the microbial cell is the governance of transmembrane ionic gradients. For example, DUF368-containing proteins appear to contribute most in alkaline conditions where the PMF is low[38], suggesting that if C55-P translocation is energy-dependent, DUF368-mediated transport may depend on the SMF or another non-PMF energy gradient. It is also possible that protonation states of C55-P, which are probably pH-dependent, determine the maximally active translocase or potential contribution of spontaneous C55-P scrambling. Characterizing the potential energetic reliance and directionality of multiple C55-P translocases will be an important topic of future study.

The conditionality of DUF368-containing and DedA family proteins, and perhaps additional as-yet-unidentified translocases, could support C55-P

flux in diverse microbial niches (Fig. 5h). The sensitivity of Δ*vca0040 V. cholerae* to increased sodium ion concentration and the apparent enrichment of DUF368 in other known halophiles such as *S. aureus* and the archaeal class Halobacteria (including *H. volcanii*) suggests that regardless of their energetic input, DUF368-containing proteins may facilitate microbial adaptation to high-salt environments (Supplementary Table 1). Specifically, the sodium and pH-dependence of the contribution of VCA0040 to *V. cholerae* physiology may reflect the robust capacity of the cholera pathogen to colonize the human small intestine, where alkaline conditions and millimolar-scale sodium concentrations are likely to occur[39]. Alongside the variable conservation of redundant translocases (Extended Data Fig. 10a and Supplementary Table 9), additional inputs and adaptation mechanisms likely also shape how C55-P recycling contributes to microbial fitness. For example, although *vca0040* is required only in alkaline conditions in *V. cholerae*, this gene is essential in the related pathogen *Vibrio parahaemolyticus* even at neutral pH, despite the conservation of *yghB* and suppressor loci, suggesting that the pH setpoint at which DUF368 family activity is dominant is not universal[40] (Extended Data Fig. 10b). Although DUF368-containing proteins—which have been renamed polyprenyl phosphate transporter (PopT) proteins in concurrently reported work[41]—are restricted to bacteria and archaea, DedA family members are widely present in eukaryotes, including humans. Thus, our findings may affect the understanding of polyprenyl phosphate (for example, dolichol phosphate) translocation in all kingdoms of life.

Methods

Bacterial strains, media and growth conditions

All *V. cholerae* strains used in this study are derivatives of HaitiWT, a spontaneous streptomycin-resistant variant of a clinical isolate from the 2010 Haiti cholera epidemic[42]. All *S. aureus* strains used in this study are derivatives of HG003 (itself a derivative of NCTC8325). Strain and plasmid information is listed in Supplementary Table 10. All constructed strains were verified by Sanger sequencing of the targeted locus (Genewiz).

The following media were used in this study: lysogeny broth (LB) Miller (10 g l^{-1} NaCl) (BD Biosciences, USA), M9 minimal media (12.8 g l^{-1} $Na_2HPO_4 \cdot 7H_2O$, 3 g l^{-1} KH_2PO_4, 0.5 g l^{-1} NaCl, 1 g l^{-1} NH_4Cl, 1 mM $MgSO_4$ and 10 μM $CaCl_2$) (lab-made), M63 minimal media (2 g l^{-1} $(NH_4)_2SO_4$, 13.6 g l^{-1} KH_2PO_4, 0.5 mg l^{-1} $FeSO_4 \cdot 7H_2O$ and 1 mM $MgSO_4$) (US Biologicals), TSB and tryptic soy agar (TSA) (BD Biosciences). M9 was supplemented with 0.4% glucose and pH-adjusted with NaOH. M63 was supplemented with 2% glucose and pH-adjusted with KOH. In general, LB was buffered with 50 mM Na_2HPO_4 or 100 mM Tris and pH-adjusted with NaOH or HCl, and TSB was buffered with 100 mM bicine. Variations in buffering are noted in figure legends. Plates were used at 1.5% final agar concentration. Where necessary, the following antibiotics or supplements were used: streptomycin (200 μg ml^{-1}), carbenicillin (50 μg ml^{-1}), kanamycin (50 μg ml^{-1}), neomycin (50 ug ml^{-1}), targocil (10 μg ml^{-1}), erythromycin (10 μg ml^{-1}), diaminopimelic acid (DAP, 0.3 mM), 5-bromo-4-chloro-3-indolyl β-d-galactopyranoside (X-gal, 60 μg ml^{-1}), arabinose (0.2%) or IPTG (1 mM).

For routine culture, *V. cholerae* were grown in non-buffered LB at 37 °C and *S. aureus* were grown in non-buffered TSB at 37 °C. Cultures for spent supernatant analysis in *V. cholerae* were obtained by subculturing 37 °C-grown overnight cultures 1:1,000 in fresh media and growing for 24 h at 30 °C. To collect spent supernatants, cultures were centrifuged for 10 min at 5,000*g* at 4 °C. Supernatants were transferred to a new tube, re-centrifuged, and sterile filtered with 0.22 μm syringe filters into new tubes. Supernatants were stored at 4 °C for future use. Cultures used for pH quantification were centrifuged for 10 min at 5,000*g* before measurement. All pH measurements, including for media and buffer formulation, were performed with a pH meter (Thermo ORION) freshly calibrated to two appropriate pH standards from 4, 7 and 10.

Bioinformatic methods

Information on the loci, genomes, and accession numbers for genes and proteins from this study is available in Supplementary Table 10.

Genomic locus correspondence in *V. cholerae*

During our study, which mainly used a contemporary pandemic strain of *V. cholerae* (HaitiWT, also known as KW3, assembly accession GCA_001318185.1), we noted several annotation inconsistencies with genes ostensibly conserved in the reference shotgun assembly (GCA_000006745.1) of N16961 *V. cholerae*. Compared to newer long-read assemblies (GCA_003063785.1 and GCA_900205735.1), the original assembly of N16961 has ~50 more ORFs annotated as pseudogenes due to frameshifts that are not included in coding sequence tables. Three of these were directly related to our study: (1) *lacZ*, a widely used reporter gene (VC2338), (2) *secF2* (VCA0692), and (3) *yghB* (VCA0534). *lacZ* is known to be functional in N16961, and re-sequencing of our laboratory N16961 stock and examination of newer N16961 assemblies confirmed that *secF2* and *yghB* are in fact intact in this strain and lack the predicted frameshift. The 'pseudogene' annotation of *secF2* led to its recent usage as a safe-harbour locus for genome editing in *V. cholerae*[43], and caution is probably required for this method. To resolve the issues of difference between strain genomes, a batch BLAST dictionary assigning putative VC locus tags (the conventional *V. cholerae* gene naming system) to annotated HaitiWT loci is given in Supplementary Table 11.

Phylogenetic and domain architecture analysis

To identify sequences with annotated DUF368 domains, Annotree[44] was used with the Pfam identifier PF04018 at a cut-off e-value of 1×10^{-30}. The results from this search are included in Supplementary Table 1. Taxa with '_X' names were manually collated into a single group for ease of analysis. In a related effort, to identify annotation-independent homologues of *vca0040*, we used a HMMER search with the VCA0040 sequence from wild-type *V. cholerae* (Supplementary Table 2). To analyse the relative distributions of PF04018, PF09335 (DedA), and PF02673 (BacA), we used Annotree at a cut-off e-value of 1×10^{-15}. Differential species lists with all possible conservation combinations were visualized with a Venn diagram generator (https://bioinformatics.psb.ugent.be/webtools/Venn/) and are provided in Supplementary Table 9.

To assess domain architecture, we used InterProScan (https://www.ebi.ac.uk/interpro) with DUF368 (PF04018) or DedA (PF09335) as queries and manually examined identified domain structures. Sequences with domain architectures relevant to this study were exported and collated in Supplementary Table 3.

Structural modelling

Structural prediction of DUF368 and DedA family proteins in this study was performed with AlphaFold2 on the CoLabFold publicly accessible interface[45]. Sequences were modelled as monomers using mmseqs2 for multiple sequence alignment. Structures were ranked by pLDDT and the top-ranked structures were visualized with ChimeraX (UCSF).

Cloning, vectors and strain construction

V. cholerae

Cloning of expression vectors and deletion plasmids was performed by standard isothermal assembly techniques with 25-bp overhangs on each fragment (HiFi DNA Assembly Kit, NEB). *V. cholerae* mutants were generated by allelic exchange as previously described with the suicide vector pCVD442 bearing 500–700 bp upstream and downstream homology arms of the targeted locus[46,47]. Either SM10λpir or MFDλpir *E. coli* (a DAP auxotroph) were used as the donor strain with identical conjugation conditions apart from DAP supplementation. Single crossover transconjugants were isolated by selective plating on LB + streptomycin/carbenicillin and passaged in 10% sucrose overnight at room temperature to select for double crossover events. Cells were plated on LB + streptomycin, re-patched onto LB + streptomycin and LB + streptomycin/carbenicillin, and streptomycin- and carbenicillin-resistant colonies were screened by colony PCR for successful deletion. For deletions involving *vca0040*, *vca0040* was always deleted last and at least two clones per strain were stocked and verified for the correct phenotype to guard against spontaneous suppressor formation. For native *vca0040* expression from the ectopic chromosomal site, the full coding sequence along with 350

bp upstream was cloned to include the native promoter and inserted at a known neutral genomic location in *V. cholerae*[48]. For expression of *SAOUHSC_00846* from the *vca0040* locus in Δ*vca0040 V. cholerae*, we used allelic exchange to replace the *vca0040* coding sequence with a *SAOUHSC_00846* sequence codon-optimized for *V. cholerae*. For depletion of *yghB* from Δ*vca0040 V. cholerae*, we used pAM299, a suicide vector which replaces the native allele of targeted gene with an arabinose-inducible copy[49]. For overexpression studies, we used pBAD18 for plasmid-based arabinose induction[50].

S. aureus

S. aureus Δ*SAOUHSC_00846* was generated with a one-step allelic exchange using the pTarKO plasmid bearing 1,000 bp upstream and downstream homology arms of the targeted locus flanking a kanamycin cassette as previously described[51]. In brief, the assembled pTarKO plasmid was electroporated into electrocompetent *S. aureus* RN4220 $tarO_{off}$ and selected for double crossovers on TSB with kanamycin, neomycin and targocil. The kanamycin insertion in *SAOUHSC_00846* was then transduced to *S. aureus* HG003 with phage phi85. An identical protocol was used to generate Δ*SAOUHSC_02816*. For Δ*SAOUHSC_0901*, the strain NE1150, which carries an erythromycin-resistance-marked transposon insertion in *SAOUHSC_00901*[52], was used as the donor for transduction into HG003. For overexpression studies, we used pLOW for plasmid-based IPTG induction[53]. Assembled plasmids were electroporated into RN4220 wild type and transduced to *S. aureus* HG003 Δ*SAOUHSC_00846* with phage phi85.

E. coli lptD4213 arabinose suppressor selection

E. coli MC4100 and its derivative *lptD4213* are sensitive to arabinose due to the araD139 mutation. To enable the use of the pBAD arabinose expression system, spontaneous arabinose-resistant mutants of MC4100 *lptd4213* were isolated as previously described[54]. In brief, overnight cultures of *lptD4213 E. coli* were plated (~10 µl) on LB + 0.4% arabinose plates and incubated at 37 °C. Putative suppressor colonies were purified on a new LB + 0.4%

arabinose plate. Suppressor colonies were then screened for growth on LB + 0.4% arabinose and no growth on M9 minimal agar + 0.2% arabinose, in order to select for strains that are resistant to arabinose, but unable to use arabinose as carbon source (that is, *ara*$^{-/r}$, but not *ara*$^{+}$ colonies). To transform *E. coli* with various vectors, competent cells were heat shocked according to a standard protocol and selected on LB supplemented with the appropriate antibiotic(s).

V. parahaemolyticus

V. parahaemolyticus RIMD2210633 mutants were constructed with a similar allelic exchange system to *V. cholerae* as previously described, with the suicide vector pDM4[40].

Growth assays

Growth curves in liquid medium

For automated A_{600} measurements, 1 ml of a saturated 37 °C LB overnight *V. cholerae* culture was washed once with 1 ml fresh media (specific to the experiment) and resuspended in 1 ml fresh media. Resuspended bacteria were diluted to a starting dilution of 1:4,000 by performing a 1:100 dilution into 1 ml fresh media and a subsequent 1:40 dilution into 195 µl of specific media aliquoted into sterile 96-well plates (Corning). At least three technical replicates per strain and media condition were run per plate along with at least three blank media wells. Growth curves were performed in a BioTek Epoch2 spectrophotometer with shaking and A_{600} readings were taken every 10 min for 20–24 h. Gen5 version 3.08 (BioTek) was used for data collection. Data for each condition were averaged across technical and biological replicates corrected against the baseline blank A_{600} values.

Overexpression vector plate dilution assays on solid medium

V. cholerae strains were grown overnight at 37 °C in LB with addition of 50 µg ml^{-1} carbenicillin for strains with pBAD18. The cultures were diluted

1:100 into 5 ml of LB (supplemented with 50 μg ml^{-1} carbenicillin and 0.2% arabinose for strains with pBAD18) and grown at 37 °C to an OD_{600} of ~0.7. Cells were normalized to an A_{600} of 0.1 and then serially tenfold diluted 6 times. Five microlitres of the dilution series were plated on LB agar 100 mM Tris pH 9 (with 0.2% arabinose, 0.2% glucose, or no sugar added) and incubated at 30 °C for 18–24 h.

S. aureus strains were grown overnight at 37 °C in TSB with addition of 10 μg ml^{-1} erythromycin for strains with pLOW. The cultures were diluted 1:100 into 5 ml of TSB (supplemented with 10 μg ml^{-1} erythromycin and 1 mM IPTG for strains with pLOW) and grown at 37 °C to an A_{600} of ~0.7. Cells were normalized to an A_{600} of 0.1 and then serially tenfold diluted 6 times. Five microlitres of the dilution series were plated on TSA 100 mM Tris pH 9 (with or without 1 mM IPTG) and incubated at 37 °C for 18–24 h.

Live and single timepoint phase-contrast microscopy

For single timepoint imaging of live cells, samples from the indicated cultures were concentrated as necessary and immobilized on 0.8% agarose pads in sterile PBS on glass slides (Gene Frames, Thermo) and dried before coverslip placement. For time-lapse imaging of live cells, samples were spotted on 0.8% agarose pads in sterile LB before imaging in a temperature-controlled chamber at a rate of 1 frame per minute for 3 h of growth. Cells were imaged with a Nikon Eclipse Ti microscope equipped with an Andor NeoZyla camera and a 100× oil phase 3 1.4-numerical-aperture (NA) objective using NIS Elements AR software version 4.6 (Nikon). Image analysis was performed with ImageJ version 1.53. Images in manuscript figures are representative of at least 10 fields of view (FOVs) from the same sample (>200 cells) and multiple independent replicate cultures as indicated.

Sphere formation and incubation assays

To induce sphere formation, *V. cholerae* 37 °C overnight cultures (where Δ*vca0040* cells are still largely rod-shaped) were back-diluted 1:100 into the indicated supernatant or fresh media and grown for 4 h shaking at 200 rpm at 30 °C. Then, cells were imaged as described above. For the d/l-Ala treatment

assay, overnight cultures were expanded in fresh LB for 90 min prior to spike-in of the amino acid for 1 h.

MIC assays

V. cholerae

To quantify MICs for various antimicrobial agents, 37 °C overnight cultures of *V. cholerae* were diluted 1:100,000 in fresh media. Diluted cultures were used to inoculate 96-well plates containing 12 twofold dilutions of the indicated agent in LB medium at a ratio of 50 µl culture:50 µl medium. Four technical replicates were performed per strain per dilution. Plates were incubated for 24 h at 37 °C and MIC values were read as the first dilution where no turbidity was observed. For repeat assays, MICs were performed with independent overnight cultures.

S. aureus and *E. coli*

Overnight cultures (at 37 °C) of *S. aureus* or *E. coli* were diluted to A_{600} = 0.01 and then further diluted 1:100 in fresh TSB media. Diluted cultures were added to 96-well plates containing twofold dilutions of the indicated agent in TSB medium or TSB medium 5 mM Tris pH 8.5 at a ratio of 75 µl culture:75 µl media. Plates were incubated for 24 h at 30 °C (37 °C for *E. coli*) with shaking. MIC values were read off as the first dilution where no turbidity was observed.

For *E. coli lptD4213* vector overexpression MICs, overnight cultures were diluted 1:100 into non-buffered LB + carbenicillin + 0.2% arabinose. The culture was outgrown for 1.5–2 h and diluted to A_{600} = 0.01. This was further diluted to A_{600} = 0.0001 and 75 µl of the culture was added to 75 µl of serial dilution of the drug. Diluted cultures were added to 96-well plates containing 2-fold dilutions of the indicated agent in is LB 5 mM Tris pH 8.5, 1 mM $CaCl_2$, with 0.2% glucose or 0.2% arabinose at a ratio of 75 µl culture: 75 µl media. Plates were incubated and read as described above.

Peptidoglycan characterization

Preparation of *V. cholerae* and *S. aureus* sacculi

For *V. cholerae*, strains were grown overnight at 37 °C in LB + 200 μg ml^{-1} streptomycin. For each sample, 500 μl of overnight culture were collected and centrifuged at 5,000*g* for 5 min, washed once with 500 μl of the corresponding media, and resuspended in 500 μl of the corresponding media. This culture was then added to 50 ml of M63 media (pH 7 with 100 mM NaCl, pH 8 with 100 mM NaCl or pH 8 with 0 mM NaCl) supplemented with 2% glucose and 200 μg ml^{-1} streptomycin and incubated at 30 °C for approximately 6 h until the A_{600} reached ~0.5. Bacteria were collected by centrifugation at 6,000*g* for 10 min at 4 °C and resuspended in 1.5 ml of 1× PBS. The resuspension was added dropwise into 1.5 ml of boiling 5% SDS solution. The mixture was boiled for 1 h and stirred for a further 2 h after the heat was turned off.

For *S. aureus*, strains were grown overnight at 37 °C in 3 ml TSB. 500 μl of the overnight culture was added to 100 ml of TSB without pH adjustment or TSB + 100 mM bicine pH 8.5 and incubated at 37 °C for ~2 h until the A_{600} reached 0.5. The bacterial cells were collected by centrifugation at 5,000*g* for 10 min at 4 °C and resuspended in 1.5 ml of 1× PBS. Resuspended samples were boiled as for *V. cholerae*.

Total peptidoglycan and cross-linking quantification from sacculi samples

Peptidoglycan was extracted from boiled samples as described previously for Gram-negative and Gram-positive bacteria[55]. Once boiled, cell wall material was pelleted by ultracentrifugation and washed with water. Clean sacculi were digested with muramidase (100 μg ml^{-1}) and soluble muropeptides reduced using 0.5 M sodium borate pH 9.5 and 10 mg ml^{-1} sodium borohydride. The pH of the samples was then adjusted to 3.5 with phosphoric acid. UPLC analyses were performed on a Waters UPLC system equipped with an ACQUITY UPLC BEH C18 Column, 130 Å, 1.7 μm, 2.1 mm × 150 mm (Waters Corporation) and identified by $A_{204\ nm}$. Muropeptides were separated using a linear gradient from buffer A (0.1% formic acid in water) to buffer B (0.1% formic acid in acetonitrile). Identification of

individual peaks was assigned by comparison of the retention times and profiles to validated chromatograms. The relative amount of each muropeptide was calculated by dividing the peak area of a muropeptide by the total area of the chromatogram. The abundance of peptidoglycan (total peptidoglycan) was assessed by normalizing the total area of the chromatogram to the A_{600}. The degree of cross-linking was calculated as described previously[56].

Intracellular UDP-M5 quantification

For *V. cholerae*, strains were grown overnight at 37 °C in LB + 200 µg ml^{-1} streptomycin. For each sample, 500 µl of overnight culture were collected and centrifuged at 5,000*g* for 5 min, washed once with 500 µl of the corresponding media, and resuspended in 500 µl of the corresponding media. Eighty microlitres of this resuspension was then added to 8 ml of M63 media (pH 7 with 100 mM NaCl, pH 8 with 100 mM NaCl or pH 8 with 0 mM NaCl) supplemented with 2% glucose and 200 µg ml^{-1} streptomycin and incubated at 30 °C for approximately 6 h until the A_{600} reached ~0.5. Bacteria cells were collected by centrifugation at 5,000*g* for 10 min at 4 °C, washed twice with 1 ml of ice cold 0.9% NaCl, and resuspended in 200 µl of milliQ water. The resuspension was boiled for 30 min, centrifuged at 20,000*g* for 15 min, and the supernatant was filtered and analysed.

For *S. aureus*, strains were grown overnight at 37 °C in 3 ml TSB. Eighty microlitres of the overnight culture was added to 8 ml of TSB without pH adjustment or TSB + 100 mM bicine pH 8.5 and incubated at 37 °C for ~2 h until the A_{600} reached ~0.5. Cultures were then processed as for *V. cholerae*.

Quantification of soluble UDP-M5 muropeptide levels by liquid chromatography–mass spectrometry (LC–MS) of filtered supernatants was performed as previously described[57]. Detection and characterization of soluble muropeptides by LC–MS was performed on an UPLC system interfaced with a Xevo G2/XS Q-TOF mass spectrometer (Waters Corporation) using previously reported conditions[57]. UDP-M5 levels were quantified by integrating peak areas from extracted ion chromatograms

(EICs) of the corresponding *m*/*z* value. UDP-M5 abundance was normalized to culture A_{600} as for total peptidoglycan.

Quantification of C55-OH and C55-P in bacterial membrane lipid extracts

Preparation of *S. aureus* or *V. cholerae* membranes

S. aureus or *V. cholerae* cells were grown overnight at 37 °C in 10 ml TSB or LB pH 7, respectively. Cells were washed, collected by centrifugation, resuspended in the experimental medium (for example, TSB or LB at a given pH) and inoculated at a 1:200 dilution into 200 ml of the experimental medium. Once an A_{600} of 0.6 was reached, cells were collected by centrifugation, resuspended in 10 ml of phosphate-buffered saline (PBS), and disrupted in a French press. Cell debris was discarded, and the supernatant centrifuged in a Beckman Optima Max TL ultracentrifuge for 15 min at 270,000*g*. Lipids from the membrane pellets were then extracted as described below.

Extraction of lipids from membrane fractions

S. aureus and *V. cholerae* membrane lipids were extracted largely as previously described[58]. Membrane pellets were resuspended in 500 µl of PBS. Then, 1.25 ml of methanol and 625 µl of chloroform were added. The suspension was vortexed for 2 min at room temperature, and the homogenates were centrifuged at 7,100*g* for 10 min at 4 °C. 625 µl of chloroform and 625 µl of PBS were added to the supernatants and they were vortexed and centrifuged at 7,100*g* for 10 min at 4 °C to separate the chloroform phase from the PBS–methanol phase. The chloroform phase was then isolated and vacuum dried. Dried pellets containing the purified membrane lipids were resuspended in 300 µl of the UPLC mobile phase solvents (50% H_2O with 0.1 % formic acid + 50% isopropanol with 0.1 % formic acid). To ensure that this method is capable of extracting C55-OH and C55-P, 200 nmol of each lipid standard (Larodan) was added to fresh purified membrane samples before subjecting them to the extraction protocol.

Detection of C55-OH and C55-P by UPLC

Lipids extracts resuspended in UPLC mobile phase solvents were analysed by UPLC on a reverse-phase C18 column Kinetex C18 UPLC Column 1.7 µm particle size, 100 Å pore size, 50 × 2.1 mm. Separation of the lipids was accomplished using a gradient of H_2O with 0.1% formic acid for solvent A and isopropanol with 0.1 % formic acid for solvent B, a flow rate of 0.4 ml/min with a linear gradient over 8 min, column temperature 60 °C and a wavelength of 210 nm. Identification of C55-OH and C55-P was based on the retention time of standards and their abundances were calculated relative to the total peak area of each chromatogram. For calculating mutant-to-wild type ratios, samples from independent cultures grown on the same day were randomly paired. Empower3 chromatography data software (Waters) was used for UPLC data collection and analysis.

Ampho-FL synthesis and staining

To synthesize ampho-FL, 5 mg of amphomycin (Cayman Chemical) was dissolved in 200 µl of dimethyl formamide (DMF) and combined with 7 µl of triethylamine. Separately, 3 mg of fluorescein-$C_{5,6}$-NHS (Thermo Fisher, USA) was dissolved in 200 µl of DMF and then added to the amphomycin solution. The reaction mixture was stirred at room temperature in the dark for 24 h. The solution was diluted in one equal volume of DMSO and purified by reverse-phase HPLC (Agilent 1260 Infinity) using a C18 stationary phase column (Luna 5 µM C18(2) 100 Å, 250 × 10 mm). HPLC conditions were as follows—phase A: water (0.1% formic acid); phase B: acetonitrile (0.1% formic acid). Phase B: 0–2 min, 50%; 2–15 min, linear gradient 50%–100%, 15–17 min, 100%. The wavelength of the detector was set at 254 nm. The flow rate was 4.7 ml min^{-1}. The mixture of conjugated fluorescein-$C_{5,6}$ products eluted at 9 min. HPLC eluates were collected in a 50 ml round-bottom flask, concentrated by rotatory evaporation, transferred with DMSO to a tared microcentrifuge tube and lyophilized, resulting in 1.7 mg of ampho-FL (M+1 = 1,649.23).

For ampho-FL labelling, overnight cultures of *S. aureus* HG003 wild type or Δ*SAOUHSC_0084*6 were diluted 1:100 in fresh TSB media buffered at pH 6 (100 mM MES), 7 (100 mM Tris), or 8.5 (100 mM bicine). The new cultures

were grown to $A_{600} = 0.5$. 1 ml of the culture was centrifuged at 1,000*g* for 5 min. The pellet was resuspended in 500 µl of 1× Tris-buffered saline (TBS) pH 9. Ninety-eight microlitres of the mixture was transferred to a new tube and combined with 2 µl of 10 mg ml^{-1} ampho-FL conjugate in DMSO. The mixture was incubated in the dark at room temperature for 10 min, washed three times with 500 µl of 1× TBS pH 9, and resuspended in 100 µl of 1× TBS pH 9. Ten microlitres bacteria was then added to a 0.8% agarose pad in TBS pH 9 containing 1 µg ml^{-1} propidium iodide and imaged as described above. To avoid bias in FOV selection, the pad was scanned with the FITC channel off to select FOVs based solely on cell density. Bacterial cells were imaged using DIC at a 30 ms exposure and ampho-FL signal was captured with the FITC filter at a 1 s exposure. Fluorescence intensity of ampho-FL was quantified using ImageJ version 1.53. FITC signal profiles were generated by drawing bisecting lines through dividing, propidium iodide-negative cells. The FITC signal was recorded at the background to the left of the cells (A), at the left wall peak (B), at the middle walls (C), at the right wall peak (D), and at the background to the right of the cells (E). The average signal was calculated by $(B + C + D)/4 - (A + E)/2$.

Transposon-insertion sequencing

Generation of transposon libraries in the Δ*vca0040* and Δ*yghB* backgrounds was performed as previously described for HaitiWT *V. cholerae*[15]. Strains were conjugated to SM10λpir *E. coli* bearing the donor transposon vector pSC189. Due to the moderate conjugation defect of Δ*vca0040*, we used a 5× concentration of the conjugation reactions compared to the other libraries. Reactions were plated on 245 mm^2 LB + streptomycin/kanamycin agar plates to isolate *V. cholerae* transconjugants and incubated overnight at 30 °C. Two independent Δ*vca0040* and Δ*yghB* libraries were generated, consisting of approximately 200,000 colonies each and stocked at an A_{600} of ~10 in LB + 25% glycerol. For synthetic transposon-insertion sequencing analyses, a frozen aliquot of each library was thawed and used for genomic DNA extraction with the Wizard kit (Promega). DNA libraries were prepared in an identical manner to previous transposon-insertion sequencing experiments from our group and sequenced on an in-house MiSeq platform (lllumina)[18]. Reads were trimmed, mapped, and processed as previously described with the Con-ARTIST transposon-insertion sequencing analysis

pipeline using Python version 3.0 and Matlab version 9[18]. To identify synthetic interaction loci, we used the wild type as the ‘input’ library and the mutant as the ‘output’ library during analysis.

RNA sequencing

For transcriptomic analyses, triplicate wild-type and Δ*vca0040 V. cholerae* 37 °C overnight cultures were back-diluted 1:100 into fresh LB medium and grown at 30 °C for 8 h with shaking. At 8 h, samples were checked by microscopy to ensure onset of sphere formation in the mutant samples, at which point 2 ml of each culture was spun down (5 min at 5000*g* at room temperature). RNA was extracted with Trizol reagent (Sigma-Aldrich) per the manufacturer’s instructions. Isolated RNA was then DNase-treated and re-isolated with ethanol precipitation according to a standard protocol. RNA samples were quality checked with a Bioanalyzer to confirm RIN values > 6. Library preparation and sequencing was performed by the Microbial Genome Sequencing Center (Pittsburgh). RNA-sequencing analysis was performed largely as described[59]. Reads were mapped to the *V. cho\lerae* KW3 genome (NCBI assembly GCA_001318185.1) with Bowtie2 (version 2.4.1)[60]. A read matrix for each sample was then generated with featureCounts from Rsubread (version 2.4.3)[61]. Read matrices were then analysed with DESeq2 (version 1.30.1) using default settings in R (version 4.0.3) to identify differentially expressed genes. Data shrinkage was performed with ashr[62].

Multiplexed tandem mass tag proteomics

For proteomic analyses, triplicate 37 °C wild-type and Δ*secDF1 V. cholerae* overnight cultures were grown as described in ‘RNA sequencing’. At 8 h of growth, whole-cell pellet (WCP) samples were prepared by centrifuging 1 ml of cells for 5 min at 5,000*g* at room temperature, washing once in fresh LB, flash frozen and kept at −80 °C. Mass spectrometry analysis was performed by the Thermo Fisher Center for Multiplexed Proteomics (TCMP) at Harvard Medical School according to standard protocols. WCP samples were subjected to a total proteomics workflow with fractionation. Pelleted cells were first lysed in 8 M urea, 200 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS) and 1% SDS with phosphatase and

protease inhibitors. Then, samples were reduced with DTT and alkylated with iodoacetamide. Alkylated proteins were precipitated with methanol/chloroform and resuspended in 200 mM EPPS pH 8 and digested sequentially with 1:50 LysC and 1:100 trypsin. Peptides were labelled with TMT16 reagents, pooled, and fractionated by a basic reverse-phase protocol into 12 fractions. Fractions were then dried, cleaned on a C18-packed stage tip, and eluted into an mass spectrometry sample vial for analysis. Samples were resuspended in 5% acetonitrile/5% formic acid and analysed by LC–MS3 on an Orbitrap Lumos mass spectrometer using a 180 min MS3 method. Peptides were detected (MS1) and quantified (MS3) in the Orbitrap, and sequenced (MS2) in the ion trap. MS2 spectra were searched with the COMET algorithm against the *V. cholerae* KW3 proteome, its reversed complement, and known contaminants. Spectral matches were filtered to a 1% false-discovery rate using the target-decoy strategy combined with linear discriminant analysis. Proteins were quantified from peptides with a summed signal to noise threshold of >150 and isolation specificity of >0.5.

Spontaneous suppressor isolation and sequencing

V. cholerae

To ensure independent isolation of spontaneous suppressors, the Δ*vca0040* strain was re-derived 11 times from different conjugation reactions. From each colony PCR- and cell shape defect-verified Δ*vca0040* clone, colonies with mutant morphology (small and completely blue) on LB + streptomycin/X-gal plates were re-streaked onto new LB + streptomycin/X-gal plates and grown for 36–48 h at 37 °C. This process was repeated once. From the tertiary plates, a single colony with wild-type morphology (large with a white halo) was re-streaked to confirm a stable suppressor phenotype. Suppressors were checked by microscopy of 30 °C overnight cultures to confirm cell shape defect reversion and verified by colony PCR to confirm the absence of *vca0040* from their genome. Validated suppressors were grown overnight, and genomic DNA was extracted as described above. Library preparation and sequencing were either performed in-house or outsourced to the Microbial Genome Sequencing Center (Pittsburgh). For in-house whole-genome sequencing, gDNA was tagmented and barcoded with the Nextera XT library preparation kit (Illumina), quality checked by

BioAnalyzer and sequenced on a MiSeq platform to at least 20–50× depth of the *V. cholerae* genome (~4 Mbp). Genome assembly and variant identification was performed with CLC Genomics Workbench 12 (Qiagen, Germany). Variants were filtered against an assembled HaitiWT isolate re-sequenced with the same workflow. Mutations were assigned as suppressors if they were present in >90% of reads and did not occur in a known poorly mapping or highly repetitive region such as a tRNA locus. For the secondary screen in *secDF2*- Δ*vca0040* bacteria, the Δ*vca0040* deletion was re-derived three independent times in the Δ*secD2* or Δ*secF2* background. Suppressors were isolated and sequenced identically to those in the parental Δ*vca0040* background.

S. aureus

Multiple 2 ml cultures of *S. aureus* HG003 Δ*SAOUHSC_00846* were grown in TSB overnight at 37 °C. One-hundred and ninety microlitres from each independent overnight culture was then separately plated on TSA 100 mM Tris pH 9 plates (where the mutant is entirely inhibited for growth) and grown at 37 °C for 24 h. Suppressors were confirmed by re-streaking on TSA 100 mM Tris pH 9 plates. Validated suppressors were grown overnight, and genomic DNA was extracted as described above. Library preparation, sequencing, and variant identification were performed as described above for *V. cholerae*.

Infant rabbit *V. cholerae* infections and caecal fluid imaging

Infant rabbit oral infections with *V. cholerae* were performed as described previously[15]. One- to two-day-old New Zealand White rabbits (Charles River Laboratories) were housed with their dams for the duration of the experiment in a temperature- and humidity-controlled facility with 12-h light/dark cycles (16–22 °C, 50% humidity). Inocula were prepared by centrifuging a late exponential phase culture (A_{600} 0.6–0.8) *V. cholerae* and resuspending in 2.5% sodium bicarbonate. For varying the pH of the inocula, sodium bicarbonate was pH-adjusted to pH 9 or 7 with NaOH or HCl immediately before resuspension. Infected kits were were orally gavaged with 10^9 CFU *V. cholerae* in a 500 µl gavage volume, returned to their dam, and monitored for 16–20 h post-infection, when they were

sacrificed by isoflurane inhalation and intracardiac injection of 20 mEq potassium chloride. Small intestinal segments and the caecum were isolated by dissection, homogenized by bead beading in PBS, and dilutions were plated on appropriate agar plates for colony enumeration. Plates were counted after an overnight incubation at 30 °C. LOD measurements were calculated by imputing a single colony at the lowest (lower bound) or highest (higher bound) dilution where a colony could reasonably have been identified, meaning that LOD values are samples where the true value is either at most (for lower bound LODs) or at least (for higher bound LODs) the limit value. During dissection of the caecum, a 28G needle was used to extract crude caecal fluid. A sample of crude caecal fluid was taken for CFU plating and imaging, and the remainder was centrifuged at 21,000*g* for 2 min. The pellet was discarded and supernatants were frozen at −20 °C for downstream analyses. For caecal fluid incubation assays, overnight cultures of *V. cholerae* were back-diluted 1:100 in caecal fluid and grown for 4 h at 30 °C shaking before imaging.

Mouse *S. aureus* intravenous infection

S. aureus HG003 (wild type) and KanR-marked Δ*SAOUHSC_00846* were grown overnight at 37 °C. Cultures were mixed at a 1:1 ratio, combined 1:1 with 50% glycerol, and stored at frozen at −80 °C in several aliquots. For infections, an aliquot was thawed and diluted in PBS to a density of ~3×10^7 CFU ml^{-1}. One-hundred microlitres was injected into the tail vein of 8- to 9-week-old female Swiss Webster mice. Mice were monitored daily and housed in a temperature- and humidity-controlled facility with 12-h light/dark cycles (20–24 °C, 50% humidity). At days 2 and 5 post-infection, mice were sacrificed by isoflurane inhalation and cervical dislocation and the heart, lungs, right kidney, liver and spleen were collected. Organs were homogenized with stainless steel beads in PBS and plated on TSA and TSA + kanamycin (TSAK) and incubated overnight to enumerate total Δ*SAOUHSC_00846* CFU, respectively. LOD calculations were performed as for infant rabbit experiments.

Animal use statement

All animal work in this study was performed in accordance with the NIH Guide on Use of and Care for Laboratory Animals and with the approval of the Brigham and Women's Hospital IACUC (protocol no. 2016N000334 for infant rabbits and protocol no. 2016N000416 for mice).

Statistics and reproducibility

Statistical tests used and replicate information are indicated in the figure legends, relevant methods sections and the Reporting summary. Unless otherwise indicated, experiments were performed with at least three independent biological replicates and either representative results or aggregate data (mean ± s.d.) are shown. Sample sizes were not statistically predetermined. Investigators were not blinded to sample identity throughout the study, and randomization of animals in infection experiments was performed as described in the Reporting Summary. All replicates resulted in similar results. Statistical analyses were performed in Prism 9 (Graphpad) and Microsoft Excel 2020.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequence data were deposited to the NCBI Sequence Read Archive under BioProject accession codes PRJNA868324 (*V. cholerae* suppressor genome sequencing), PRJNA868332 (*V. cholerae* RNA sequencing), PRJNA877769 (*V. cholerae* transposon-insertion sequencing) and PRJNA877773 (*S. aureus* suppressor genome sequencing). Proteomics data were deposited in MassIVE under the accession code MSV000090217. All other data are freely available without restriction from the corresponding authors upon request. Source data are provided with this paper.

References

1. Egan, A. J. F., Errington, J. & Vollmer, W. Regulation of peptidoglycan synthesis and remodelling. *Nat. Rev. Microbiol.* **18**, 446–460 (2020).

2. Mueller, E. A. & Levin, P. A. Bacterial cell wall quality control during environmental stress. *mBio* https://doi.org/10.1128/mBio.02456-20 (2020).

3. Vijaranakul, U., Nadakavukaren, M. J., de Jonge, B. L., Wilkinson, B. J. & Jayaswal, R. K. Increased cell size and shortened peptidoglycan interpeptide bridge of NaCl-stressed *Staphylococcus aureus* and their reversal by glycine betaine. *J. Bacteriol.* **177**, 5116–5121 (1995).

4. Madiraju, M. V. V. S., Brunner, D. P. & Wilkinson, B. J. Effects of temperature, NaCl, and methicillin on penicillin-binding proteins, growth, peptidoglycan synthesis, and autolysis in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **31**, 1727–1733 (1987).

5. Mainardi, J. et al. Resistance to cefotaxime and peptidoglycan composition in *Enterococcus faecalis* are influenced by exogenous sodium chloride. *Microbiology* **144**, 2679–2685 (1998).

6. Pazos, M., Peters, K. & Vollmer, W. Robust peptidoglycan growth by dynamic and variable multi-protein complexes. *Curr. Opin. Microbiol.* **36**, 55–61 (2017).

7. Workman, S. D. & Strynadka, N. C. J. A slippery scaffold: synthesis and recycling of the bacterial cell wall carrier lipid. *J. Mol. Biol.* **432**, 4964–4982 (2020).

8. Sham, L.-T. et al. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* **345**, 220–222 (2014).

9. El Ghachi, M. et al. Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nat. Commun.* **9**, 1078 (2018).

10. Workman, S. D., Worrall, L. J. & Strynadka, N. C. J. Crystal structure of an intramembranal phosphatase central to bacterial cell-wall

peptidoglycan biosynthesis and lipid recycling. *Nat. Commun.* **9**, 1159 (2018).

11. Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M. & Brady, S. F. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. *Science* **376**, 991–996 (2022).

12. Häse, C. C. & Barquera, B. Role of sodium bioenergetics in *Vibrio cholerae*. *Biochim. Biophys. Acta* **1505**, 169–178 (2001).

13. Juárez, O. & Barquera, B. Insights into the mechanism of electron transfer and sodium translocation of the Na^+-pumping NADH:quinone oxidoreductase. *Biochim. Biophys. Acta* **1817**, 1823–1832 (2012).

14. Alvarez, L., Hernandez, S. B. & Cava, F. Cell wall biology of *Vibrio cholerae*. *Annu. Rev. Microbiol.* **75**, 151–174 (2021).

15. Hubbard, T. P. et al. A live vaccine rapidly protects against cholera in an infant rabbit model. *Sci. Transl. Med.* **10**, eaap8423 (2018).

16. Kamp, H. D., Patimalla-Dipali, B., Lazinski, D. W., Wallace-Gadsden, F. & Camilli, A. Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathog.* **9**, e1003800 (2013).

17. Fu, Y., Waldor, M. K. & Mekalanos, J. J. Tn-Seq analysis of *Vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe* **14**, 652–663 (2013).

18. Pritchard, J. R. et al. ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. *PLoS Genet.* **10**, e1004782 (2014).

19. Lam, H. et al. d-Amino acids govern stationary phase cell wall remodeling in bacteria. *Science* **325**, 1552–1555 (2009).

20. Touzé, T., Blanot, D. & Mengin-Lecreulx, D. Substrate specificity and membrane topology of *Escherichia coli* PgpB, an undecaprenyl pyrophosphate phosphatase. *J. Biol. Chem.* **283**, 16573–16583 (2008).

21. Schneider, T. et al. The lipopeptide antibiotic Friulimicin B inhibits cell wall biosynthesis through complex formation with bactoprenol phosphate. *Antimicrob. Agents Chemother.* **53**, 1610–1618 (2009).

22. Rubinchik, E. et al. Mechanism of action and limited cross-resistance of new lipopeptide MX-2401. *Antimicrob. Agents Chemother.* **55**, 2743–2754 (2011).

23. Tanaka, H. et al. Studies on bacterial cell wall inhibitors. II. Inhibition of peptidoglycan synthesis in vivo and in vitro by amphomycin. *Biochim. Biophys. Acta* **497**, 633–640 (1977).

24. Singh, M., Chang, J., Coffman, L. & Kim, S. J. Solid-state NMR characterization of amphomycin effects on peptidoglycan and wall teichoic acid biosyntheses in *Staphylococcus aureus*. *Sci. Rep.* **6**, 31757 (2016).

25. Campbell, J. et al. Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in *Staphylococcus aureus*. *ACS Chem. Biol.* **6**, 106–116 (2011).

26. Storm, D. R. & Strominger, J. L. Complex formation between bacitracin peptides and isoprenyl pyrophosphates. The specificity of lipid–peptide interactions. *J. Biol. Chem.* **248**, 3940–3945 (1973).

27. Sampson, B. A., Misra, R. & Benson, S. A. Identification and characterization of a new gene of *Escherichia coli* K-12 involved in outer membrane permeability. *Genetics* **122**, 491–501 (1989).

28. Baker, B. R., Ives, C. M., Bray, A., Caffrey, M. & Cochrane, S. A. Undecaprenol kinase: function, mechanism and substrate specificity of a potential antibiotic target. *Eur. J. Med. Chem.* **210**, 113062 (2021).

29. Tiyanont, K. et al. Imaging peptidoglycan biosynthesis in *Bacillus subtilis* with fluorescent antibiotics. *Proc. Natl Acad. Sci. USA* **103**, 11033–11038 (2006).

30. Doerrler, W. T., Sikdar, R., Kumar, S. & Boughner, L. A. New functions for the ancient DedA membrane protein family. *J. Bacteriol.* **195**, 3–11 (2013).

31. Okawa, F. et al. Evolution and insights into the structure and function of the DedA superfamily containing TMEM41B and VMP1. *J. Cell Sci.* **134**, jcs255877 (2021).

32. Kumar, S. & Doerrler, W. T. Members of the conserved DedA family are likely membrane transporters and are required for drug resistance in *Escherichia col*i. *Antimicrob. Agents Chemother.* **58**, 923–930 (2014).

33. Panta, P. R. & Doerrler, W. T. A *Burkholderia thailandensis* DedA family membrane protein is required for proton motive force dependent lipid a modification. *Front. Microbiol.* **11**, 618389 (2020).

34. Dong, P. et al. A UPF0118 family protein with uncharacterized function from the moderate halophile *Halobacillus andaensis* represents a novel class of Na^+ (Li^+)/H^+ antiporter. *Sci. Rep.* **7**, 45936 (2017).

35. Stamsås, G. A. et al. CozEa and CozEb play overlapping and essential roles in controlling cell division in *Staphylococcus aureus*. *Mol. Microbiol.* **109**, 615–632 (2018).

36. Ishii, E. et al. Nascent chain-monitored remodeling of the Sec machinery for salinity adaptation of marine bacteria. *Proc. Natl Acad. Sci. USA* **112**, E5513–E5522 (2015).

37. Ritchie, J. M., Rui, H., Bronson, R. T. & Waldor, M. K. Back to the future: studying cholera pathogenesis using infant rabbits. *mBio* https://doi.org/10.1128/mBio.00047-10 (2010).

38. Krulwich, T. A., Sachs, G. & Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* **9**, 330–343 (2011).

39. Molla, A. M., Rhman, M., Sarker, S. A., Sack, D. A. & Molla, A. Stool electrolyte content and purging rates in diarrhea caused by rotavirus, enterotoxigenic *E. coli* and *V. cholerae* in children. *J. Pediatr.* **98**, 835–838 (1981).

40. Hubbard, T. P. et al. Genetic analysis of *Vibrio parahaemolyticus* intestinal colonization. *Proc. Natl Acad. Sci. USA* **113**, 6283–6288 (2016).

41. Roney, I. J. & Rudner, D. Z. Two broadly conserved families of polyprenyl-phosphate transporters. *Nature* https://doi.org/10.1038/s41586-022-05587-z (2022).

42. Chin, C.-S. et al. The origin of the Haitian cholera outbreak strain. *N. Engl. J. Med.* **364**, 33–42 (2011).

43. Dalia, T. N., Chlebek, J. L. & Dalia, A. B. A modular chromosomally integrated toolkit for ectopic gene expression in Vibrio cholerae. *Sci Rep.* **10**, 15398 (2020).

44. Mendler, K. et al. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **47**, 4442–4448 (2019).

45. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

46. Sit, B. *Insights into* Vibrio cholerae *Vaccine Development and Physiology From Small Animal Models of Intestinal Colonization and Disease*. PhD thesis, Harvard University (2021).

47. Sit, B. et al. Oral immunization with a probiotic cholera vaccine induces broad protective immunity against *Vibrio cholerae* colonization and disease in mice. *PLoS Negl. Trop. Dis.* **13**, e0007417 (2019).

48. Abel, S. et al. Sequence tag-based analysis of microbial population dynamics. *Nat. Methods* **12**, 223–226 (2015).

49. Möll, A. et al. A d,d-carboxypeptidase is required for *Vibrio cholerae* halotolerance. *Environ. Microbiol.* **17**, 527–540 (2015).

50. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

51. Lee, W. et al. Antibiotic combinations that enable one-step, targeted mutagenesis of chromosomal genes. *ACS Infect. Dis.* **4**, 1007–1018 (2018).

52. Fey, P. D. et al. A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. *mBio* **4**, e00537–12 (2013).

53. Liew, A. T. F. et al. A simple plasmid-based system that allows rapid generation of tightly controlled gene expression in *Staphylococcus aureus*. *Microbiology* **157**, 666–676 (2011).

54. Englesberg, E. et al. l-Arabinose-sensitive, l-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J. Bacteriol.* **84**, 137–146 (1962).

55. Alvarez, L., Hernandez, S. B., de Pedro, M. A. & Cava, F. Ultra-sensitive, high-resolution liquid chromatography methods for the high-throughput quantitative analysis of bacterial cell wall chemistry and structure. *Methods Mol. Biol.* **1440**, 11–27 (2016).

56. Alvarez, L., Cordier, B., Van Teeffelen, S. & Cava, F. Analysis of Gram-negative bacteria peptidoglycan by ultra-performance liquid chromatography. *Bio. Protoc.* **10**, e3780. (2020).

57. Hernández, S. B., Dörr, T., Waldor, M. K. & Cava, F. Modulation of peptidoglycan synthesis by recycled cell wall tetrapeptides. *Cell Rep.* **31**, 107578 (2020).

58. Barreteau, H. et al. Quantitative high-performance liquid chromatography analysis of the pool levels of undecaprenyl phosphate and its derivatives in bacterial membranes. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 213–220 (2009).

59. Warr, A. R., Kuehl, C. J. & Waldor, M. K. Shiga toxin remodels the intestinal epithelial transcriptional response to enterohemorrhagic *Escherichia coli*. *PLoS Pathog.* **17**, e1009290 (2021).

60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

61. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

62. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).

Acknowledgements

This work was supported by The National Institutes of Health (R01AI-042347 to M.K.W. and F31AI156949 to K.H.), Howard Hughes Medical Institute (M.K.W.), National Sciences and Engineering Research Council of Canada (PGSD3-487259-2016 to B.S.), Swedish Research Council (F.C.), Knut och Alice Wallenbergs Stiftelse (F.C.), Laboratory of Molecular Infection Medicine Sweden (F.C.) and Kempe Foundation (F.C.) We thank C. Wegler in the Cava laboratory for technical assistance with C55-P analysis; members of the Waldor laboratory for comments and discussions during the study; D. Rudner for coordinating submission; the Thermo Fisher Scientific Center for Multiplexed Proteomics at Harvard Medical School, especially R. Rodrigues, for assistance with proteomics; and the Microbial Genome Sequencing Center (MiGS) for assistance with RNA sequencing and whole-genome sequencing. Schematics and model figures in Figs. 1, 3 and 5 and Extended Data Figs. 5, 7 and 9 were generated with BioRender.com. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author information

Author notes

1. Brandon Sit

   Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

2. Veerasak Srisuknimit

   Present address: Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

3. These authors contributed equally: Brandon Sit, Veerasak Srisuknimit, Emilio Bueno

Authors and Affiliations

1. Division of Infectious Diseases, Brigham and Women's Hospital, Boston, MA, USA

   Brandon Sit, Veerasak Srisuknimit, Franz G. Zingl, Karthik Hullahalli & Matthew K. Waldor

2. Department of Microbiology, Harvard Medical School, Boston, MA, USA

   Brandon Sit, Veerasak Srisuknimit, Franz G. Zingl, Karthik Hullahalli & Matthew K. Waldor

3. Laboratory for Molecular Infection Medicine Sweden (MIMS), Department of Molecular Biology, Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden

   Emilio Bueno & Felipe Cava

4. Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA, USA

   Matthew K. Waldor

5. Howard Hughes Medical Institute, Bethesda, MD, USA

Matthew K. Waldor

Contributions

B.S., V.S. and M.K.W. conceived the study. B.S., V.S. and E.B. designed experiments with supervision from F.C. and M.K.W. B.S. performed in vitro and in vivo *V. cholerae* experiments, as well as bioinformatic analyses and data visualization. V.S. performed in vitro *S. aureus* experiments, *E. coli lptd4213* experiments, and synthesized ampho-FL. E.B. performed biochemical peptidoglycan and C55-P extraction and quantification. F.G.Z. performed in vitro experiments and aided in data replication. K.H. identified *V. cholerae* spontaneous suppressors, performed the in vivo *S. aureus* experiment, and aided in data replication. All authors participated in data analysis and interpretation. B.S. and M.K.W. wrote the manuscript with input and editing from all other authors.

Corresponding authors

Correspondence to Felipe Cava or Matthew K. Waldor.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Carol Gross, Anant Menon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Identification and conservation of VCA0040.

(**a**): Infant rabbit transposon-insertion sequencing data reproduced from Hubbard *et al*. (ref. [15]) identifying *vca0040* (red) as a *V. cholerae* intestinal colonization determinant. The *tcp* operon, known to be critical for colonization, is highlighted in blue. One representative rabbit (#403) is shown. Inverse p-values were obtained by the Mann-Whitney U test without adjustments for multiple comparisons. Plotted thresholds are the same as Fig. 4a. (**b** and **c**): Domain structure, locus tags, accession numbers (**b**), and alignment (**c**) of *V. cholerae* and *S. aureus* DUF368-containing proteins. Alignment was performed with T-COFFEE. IN: predicted cytosol-facing sequence. HEL: predicted transmembrane helix sequence. OUT: predicted external-facing sequence. (**d** and **e**): Phylogenetic conservation of DUF368 domains (PF04018). (**d**): Family-level conservation (blue lines) grouped by phylum (colored arcs). Selected well-represented phyla are labeled. Blue lines are not scaled and do not provide information on the number or diversity of within-family DUF368 sequences. (**e**): Family-level breakdown of DUF368 conservation.

Extended Data Fig. 2 Structural modeling of DUF368-containing proteins.

The *V. cholerae* (left) and *S. aureus* (right) DUF368-containing proteins were modeled using AlphaFold2 and visualized with ChimeraX. (**a**): Multiple sequence alignment (MSA) coverage (left) and pLDDT (confidence score, right) per position for each protein from AlphaFold2. (**b**): Overall ribbon structures, colored from N (violet) to C (red) termini. (**c** and **d**): Electrostatic surface potential (**c**) and lipophilicity (**d**) maps generated using default ChimeraX settings. *V. cholerae* panels are reproduced from Fig. 1c,d for clarity. Electrostatic colour scale: -10 to 10 kcal/(mol·e).

Lipophilicity colour bar: relative molecular lipophilicity as calculated by ChimeraX. (**e**): MSA coverage, pLDDT plot, and ribbon and electrostatic surface views of the DUF368 protein from *Haloferax volcanii* (HVO_RS12060).

Extended Data Fig. 3 Growth and morphology of Δ*vca0040 V. cholerae*.

(**a**): CFU plating time course of WT and Δ*vca0040 V. cholerae*. Onset of stationary phase occurs after 6 h. (**b**): Phase-contrast imaging of WT, Δ*vca0040* and Δ*vca0040* + *vca0040 V. cholerae* in overnight cultures in the indicated conditions. (**c**): Phase-contrast imaging time course of the indicated strains grown in LB at 30 °C for the indicated amount of time. (**d**): Timelapse imaging of Δ*vca0040* cells immobilized on an agarose pad from a 30 °C overnight culture. Images were acquired over 3 h of growth. (**e**): Imaging of *V. cholerae* treated with normal, boiled (10 min) or filtered (<3K MW cutoff) cell-free spent supernatant for 4 h at 30 °C. (**f**): Imaging of *V. cholerae* treated with vehicle (DMSO) or 3 mM of D- or L-Ala in fresh LB for 4 h at 30 °C. **a**: mean ± SD of n = 3 independent cultures per strain. **b**, **c**, **e**, **f**: representative images of n = 3 independent cultures per strain/condition. **d**: representative timelapse of a single cell representative of >20 sphere-shaped bacteria across n = 3 separate fields of view from a single overnight culture. Scale bars: 3 (**b**,**c**,**f**) or 5 (**d**,**e**) µm.

Extended Data Fig. 4 DUF368 functional conservation and effects on PG composition and crosslinking.

(**a** and **b**): Alkaline growth of Δ*vca0040 V. cholerae* bearing pBAD18 vectors expressing (**a**) *SAOUHSC_00846* from *S. aureus* or (**b**) *HVO_RS12060* from *H. volcanii*. (**c**-**e**): Composition and crosslinking analysis in the indicated *V. cholerae* strains at the indicated pH in M63 media. Suppressor: Δ*vca0040/ΔsecDF1*. (**f** and **g**): Same as **c**-**e** for *S. aureus* grown in the indicated conditions. **a**, **b**: representative data from n = 3 cultures per strain/condition. **c**-**g**: mean ± SD from n = 3 cultures per strain/condition. Statistical analysis was performed with one-way ANOVAs with Tukey's multiple comparison tests (**c**-**e**) or unpaired student's two-tailed t-tests (**f**, **g**).

Extended Data Fig. 5 Additional characterization of bacteria lacking DUF368-containing proteins.

(**a**-**d**): RNAseq of Δ*vca0040 V. cholerae*, with schematic (**a**), confirmation of sphere formation via imaging at sample collection (**b**), principal component analysis of RNAseq data (**c**), and MA plot of analyzed sequences (**d**) (n = 3 independent cultures per strain). An arbitrary fold change (FC) cutoff of 4 was implemented during RNAseq analysis, and genes with FC values beneath the p-value threshold of 0.05 are highlighted in red. *pgpB* and *vca0040* are labeled. (**e** and **f**): UPLC traces (**e**) and quantification of C55-P relative abundance (**f**) in *V. cholerae* grown at the indicated pH. (**g**): Ampho-FL-stained *S. aureus* grown at pH 6, 7, or 8.5. (**h**): Quantification of ampho-FL staining signal at different pHs as shown in **g**. Symbols represent the mean ampho-FL intensity of n = 1–8 individual bacterial clusters in independent cultures. Single pH 8.5 cultures were used as controls as a new batch of ampho-FL was used for this experiment. **e**, **g**: representative data of n = 3 cultures per strain/condition (except for n = 1 culture in pH 8.5 in **g**). **f**, **h**: n = 3 independent cultures. Statistical analysis was performed with (**d**) Wald tests in the DESeq2 pipeline with multiple comparison adjustments or (**f**, **h**) unpaired student's two-tailed t-tests. Scale bars, 5 µm.

Extended Data Fig. 6 Genetic interactions between DedA and DUF368 proteins.

(**a**): Synthetic transposon-insertion screen in Δ*yghB V. cholerae*. P-values were obtained by the Mann-Whitney U (MWU) test without adjustment for multiple comparisons. MFC: mean fold change in read coverage between Δ*yghB* and WT. (**b**): Growth of Δ*vca0040 V. cholerae* expressing vectors with arabinose-inducible genes on regular LB (left), LB pH 9 (middle) and LB pH 9 + arabinose (right) plates. (**c**): Growth of *V. cholerae* lacking one or both of *vca0040* and/or *yghB* on LB pH 6 plates. LB was buffered with 100 mM MES and pH-adjusted with HCl. **a**: pooled data from n = 2 independent transposon libraries. **b**, **c**: representative images from n = 3 cultures per strain/condition.

Extended Data Fig. 7 Contribution of Na^{+} to Δ*vca0040* alkaline susceptibility.

(**a**): Representative colonies of WT, Δ*vca0040*, and suppressor *V. cholerae* on LB+X-gal plates. White arrows indicate suppressors. (**b**): Spontaneous suppressor mutations of Δ*vca0040 V. cholerae*. SecD1, SecF1 and PpiD are part of the *V. cholerae* protein secretion machinery. Colored text indicates suppressors from a secondary screen in Δ*secD2*/Δ*vca0040* (blue) or Δ*secF2*Δ*vca0040* (red) *V. cholerae*. Asterisk: stop codon. (**c**): Phase-contrast imaging of overnight 30 °C LB cultures of the indicated strains in the Δ*vca0040* background. Scale bars, 3 µm. (**d**): Overnight 30 °C LB culture pH of the indicated strains. (**e**): SecDF1 and SecDF2 functions in *V. cholerae*. Possible SecDF functions: coupling ion import to SecA ATPase stimulation (bottom arrow) or Sec substrate guidance into the periplasm (top arrow). (**f**): Synthetic lethality of *secD2* in Δ*secD1 V. cholerae* using the pAM299 suicide vector for Ara-dependent *secD2* expression. (**g**): Growth of the indicated strains in M63 media with added NaCl or KCl. (**h**): Waterfall plot comparing Δ*secDF1* to WT *V. cholerae* proteome ranked by protein fold change (FC). (**i**): Growth of Δ*vca0040 V. cholerae* in [Na^+]-varying (top) or [K^+]-varying (bottom) M63 media at pH 6, 7 or 8. (**j** and **k**): Growth of WT (**j**) or Δ*vca0040* (**k**) *V. cholerae* in M63 pH 8 with a range of [NaCl]. 100mM NaCl curves from **i** are replicated in **j** and **k** for comparison. (**l** and **m**): Growth of the indicated *V. cholerae* strains on non-buffered (**l**) or pH 6 (**m**) LB with regular (~170 mM) or no added NaCl. **a**: representative image of n = 11 independent suppressor strains. **c**, **f**, **l**, **m**: representative data of n = 3 cultures per strain/condition. **d**: means ± SD of n = 3 cultures per strain. **g**, **i**, **j**, **k**: mean ± SD of at least 3 technical replicates of single cultures (representative of n = 3 cultures per strain/condition). **h**: pooled data comparing n = 4 independent cultures of each strain.

Extended Data Fig. 8 Supplemental data for infant rabbit infections.

(**a**): Total CFU in mixed infections with differing inoculum pH (n = 6 each for pH 7 and 9). (**b**): Fluid accumulation ratio (FAR) in mixed infections with differing inoculum pH. (**c**): Representative small intestinal CFU plates

from single infections showing transmission of WT *V. cholerae* (greyish translucent colonies) from WT-infected rabbits to Δ*vca0040*-infected rabbits in the same litter. Note that Δ*vca0040* colonies (white arrows) are lighter and smaller than WT colonies, making them visually distinguishable and countable. (**d**): Direct imaging of GFP+ *V. cholerae* in single infection CF samples. Large GFP-bright spheres are likely Δ*vca0040* cells in the presence of transmitted, rod-shaped WT cells. Scale bars, 3 µm. **a**: geometric means analyzed with two-tailed Mann-Whitney U tests. **b**: mean ± SD analyzed with an unpaired student's two-tailed t-test. **c**, **d**: representative images from n = 3 WT and n = 7 Δ*vca0040* singly-infected rabbits.

Source data

Extended Data Fig. 9 Intravenous (IV) infections of mice with *S. aureus*.

(**a**): Schematic of IV infections and sample harvesting workflow. TSA counts indicate total *S. aureus* CFU burden, whereas TSAK counts indicate Δ*SAOUHSC_00846* burden. N = 5 mice were used per timepoint (n = 10 total). (**b**): Raw CFU counts from TSA and TSA + kanamycin (TSAK) plates. In the absence of a competitive defect, the TSAK counts should be 50% of those on the TSA plates (and thus are very close together on the $\log_{10}$-plotted graph). (**c**): Ratio of Δ*SAOUHSC_00846* (KanR) colony-forming units (CFU) to total *S. aureus* CFU at Day 2 (left) and Day 5 (right) post-infection. The dotted line indicates the mean starting proportion from n = 2 technical replicate plates of the inoculum. Open circles represent the limit of detection (LOD). Note that samples with CFU counts below the LOD were not plotted. **b**: geometric means. **c**: means ± SD.

Source data

Extended Data Fig. 10 Distribution of putative C55-P translocases in bacteria and genetic analysis of a DUF368 protein in *V. parahaemolyticus*.

(**a**): Venn diagram of Annotree species-level outputs for PFAM queries PF04018 (DUF368), PF09335 (DedA), and PF02673 (BacA/UppP). Out of

30,255 bacterial species in the Annotree database, 29,416 (97.2%) are predicted to contain at least one of the three proteins. A list of species in each Venn diagram segment is listed in Supplementary Table 9. This is likely also an overestimate of the number of species lacking any of these three domains due to sequence divergence and potentially inadequate PFAM annotation. For example, the obligate intracellular pathogen *Rickettsia rickettsii* (GCA_000018225.1) is in the triply-absent group, but upon manual curation is known to have peptidoglycan and has an annotated DedA coding sequence, perhaps reflecting a PFAM annotation lag. Interestingly, however, *Mycoplasma* species, which do not have lipopolysaccharide or peptidoglycan (and thus may not be dependent on C55-P translocation) appear to be genuinely triply-absent except for a select few members with only a DUF368-containing protein. (**b**): Growth of *V. parahaemolyticus* lacking the DUF368-containing protein VPA1624 at neutral and acidic pH. VPA1624 genetically interacts with the *V. parahaemolyticus secDF1* loci, as deletion of these suppressors originally identified in *V. cholerae* rescues the *vpa1624*-depleted strain. Ara: arabinose, Glc: glucose. Representative data of n = 3 independent cultures per strain/condition.

Supplementary information

Reporting Summary

Peer Review File

Supplementary Table 1

Bacterial and archaeal clades with PF04018-containing proteins. Annotree was used to plot and identify species with an annotated DUF368 domain. Each tab of the spreadsheet is a different level of bacterial classification (Phylum>Order>Family>Genus>Species). Archaeal clades are combined on the tab “Archaea”.

Supplementary Table 2

Homologues of VCA0040 in bacteria. HMMER was used to identify homologous sequences to VCA0040 in the Uniprot database. Each species is listed once.

Supplementary Table 3

Notable domain architectures of DUF368- and DedA-containing proteins. Sequences were obtained by searching InterProScan with the PF family designations for DUF368 (PF04018) and DedA (PF09335).

Supplementary Table 4

RNAseq of Δ*vca0040 V. cholerae* during sphere formation. For the top 40 up-regulated genes in Δ*vca0040*, manual curation of subcellular localization was performed with SignalP and pSortb. P-values were obtained through the Wald test performed by DESeq2 with adjustments for multiple comparisons.

Supplementary Table 5

MIC data for *V. cholerae, S. aureus* and *E. coli.* Data for each organism is separated into individual worksheets. Note that *S. aureus* data is split into two worksheets, one for WT versus Δ*0846* characterization ("SA") and one for WT versus Δ*0846*, Δ*2816* and Δ*0901* ("SA_dedA"). All *V. cholerae* data is under "VC" and all *E. coli lptd4213* data is under "EC_lptd4213".

Supplementary Table 6

Synthetic transposon-insertion screening in Δ*vca0040 V. cholerae.* Note that intergenic regions (IG_) and regions with <5 informative sites (possible transposon insertion sites) were not analyzed. Hits on the second tab were thresholded at an inverse MWU p-value >100 before sorting by mean $\log_2$ fold change.

Supplementary Table 7

Synthetic transposon-insertion screening in Δ*yghB V. cholerae.* Note that this dataset has not been filtered as for Supplementary Table 6.

Supplementary Table 8

Multiplexed comparative proteomics of whole-cell WT and Δ*secDF1 V. cholerae*. Fold change was calculated by dividing the average normalized relative abundance of proteins in Δ*secDF1* by that of WT.

Supplementary Table 9

Conservation of PF04018, PF09335, and PF02763 in bacterial species. Spreadsheet was generated by Annotree queries for all possible combinations of the three protein families. The "Master" sheet lists all species-level organisms with the indicated combination of domains. The "Model Species" sheet lists selected microorganisms of particular interest and can be used to look up any given species in the reference table. Species of interest should be manually confirmed for presence or absence by BLAST to guard against misannotation or non-annotation issues.

Supplementary Table 10

Strains and vectors and genomic information used in this study. Any listed strains or vectors may be requested from the lead contact (mwaldor@research.bwh.harvard.edu).

Supplementary Table 11

BLAST dictionary for assignment of putative VC names to HaitiWT *V. cholerae* loci. All coding sequences from HaitiWT were used as queries in a batch BLAST of the N16961 *V. cholerae* proteome. The top hit by e-value was selected and VC name automatically assigned. However, as batch BLAST outputs a hit regardless of confidence, manual curation is still required for specific loci when using this dictionary.

Source data

Source Data Fig. 5

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Sit, B., Srisuknimit, V., Bueno, E. *et al.* Undecaprenyl phosphate translocases confer conditional microbial fitness. *Nature* **613**, 721–728 (2023). https://doi.org/10.1038/s41586-022-05569-1

- Received: 02 March 2022
- Accepted: 17 November 2022
- Published: 30 November 2022
- Issue Date: 26 January 2023

- DOI: https://doi.org/10.1038/s41586-022-05569-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

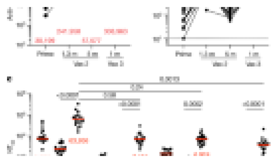
Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Two broadly conserved families of polyprenyl-phosphate transporters**
    - Ian J. Roney
    - David Z. Rudner

  Nature (2023)

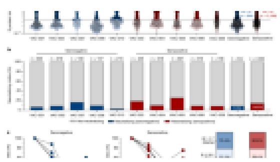
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05569-1



Increased memory B cell potency and breadth after a SARS-CoV-2 mRNA boost

21 April 2022

Frauke Muecksch, Zijun Wang, … Michel C. Nussenzweig



Hybrid immunity improves B cells and antibodies against SARS-CoV-2 variants

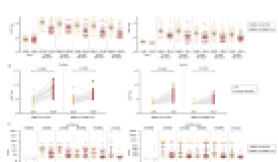
20 October 2021

Emanuele Andreano, Ida Paciello, … Rino Rappuoli

Molecular fate-mapping of serum antibody responses to repeat immunization

16 January 2023

Ariën Schiepers, Marije F. L. van 't Wout, … Gabriel D. Victora

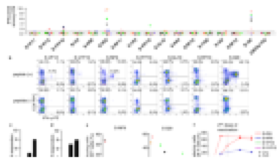


Cellular basis of enhanced humoral immunity to SARS-CoV-2 upon homologous or heterologous booster vaccination analyzed

by single-cell immune profiling

21 October 2022

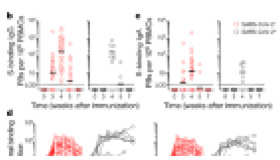
Jingwen Ai, Jingxin Guo, … Wenhong Zhang



Long-term memory CD8+ T cells specific for SARS-CoV-2 in individuals who received the BNT162b2 mRNA vaccine

06 September 2022

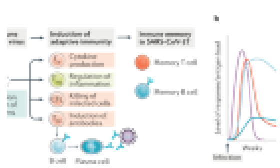
Nozomi Kuse, Yu Zhang, … Masafumi Takiguchi



SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses

28 June 2021

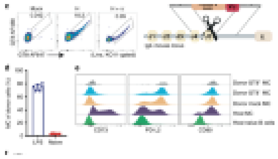
Jackson S. Turner, Jane A. O'Halloran, … Ali H. Ellebedy



Not just antibodies: B cells and T cells mediate immunity to COVID-19

24 August 2020

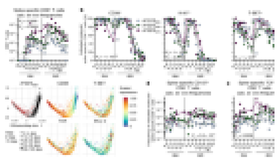
Rebecca J. Cox & Karl A. Brokstad



[Vaccine elicitation of HIV broadly neutralizing antibodies from engineered B cells]

17 November 2020

Deli Huang, Jenny Tuyet Tran, … James E. Voss



[Rapid and stable mobilization of CD8+ T cells by SARS-CoV-2 mRNA vaccine]

28 July 2021

Valerie Oberhardt, Hendrik Luxenburger, … Maike Hofmann

- Article
- Open Access
- Published: 06 December 2022

Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination

- Dennis Schaefer-Babajew,

- Zijun Wang,
- Frauke Muecksch,
- Alice Cho,
- Maximilian Loewe,
- Melissa Cipolla,
- Raphael Raspe,
- Brianna Johnson,
- Marie Canis,
- Justin DaSilva,
- Victor Ramos,
- Martina Turroja,
- Katrina G. Millard,
- Fabian Schmidt,
- Leander Witte,
- Juan Dizon,
- Irina Shimeliovich,
- Kai-Hui Yao,
- Thiago Y. Oliveira,
- Anna Gazumyan,
- Christian Gaebler,
- Paul D. Bieniasz ,
- Theodora Hatziioannou ,
- Marina Caskey &
- …
- Michel C. Nussenzweig

Show authors

Nature **volume 613**, pages 735–742 (2023)

- 12k Accesses
- 2 Citations
- 210 Altmetric
- Metrics details

Abstract

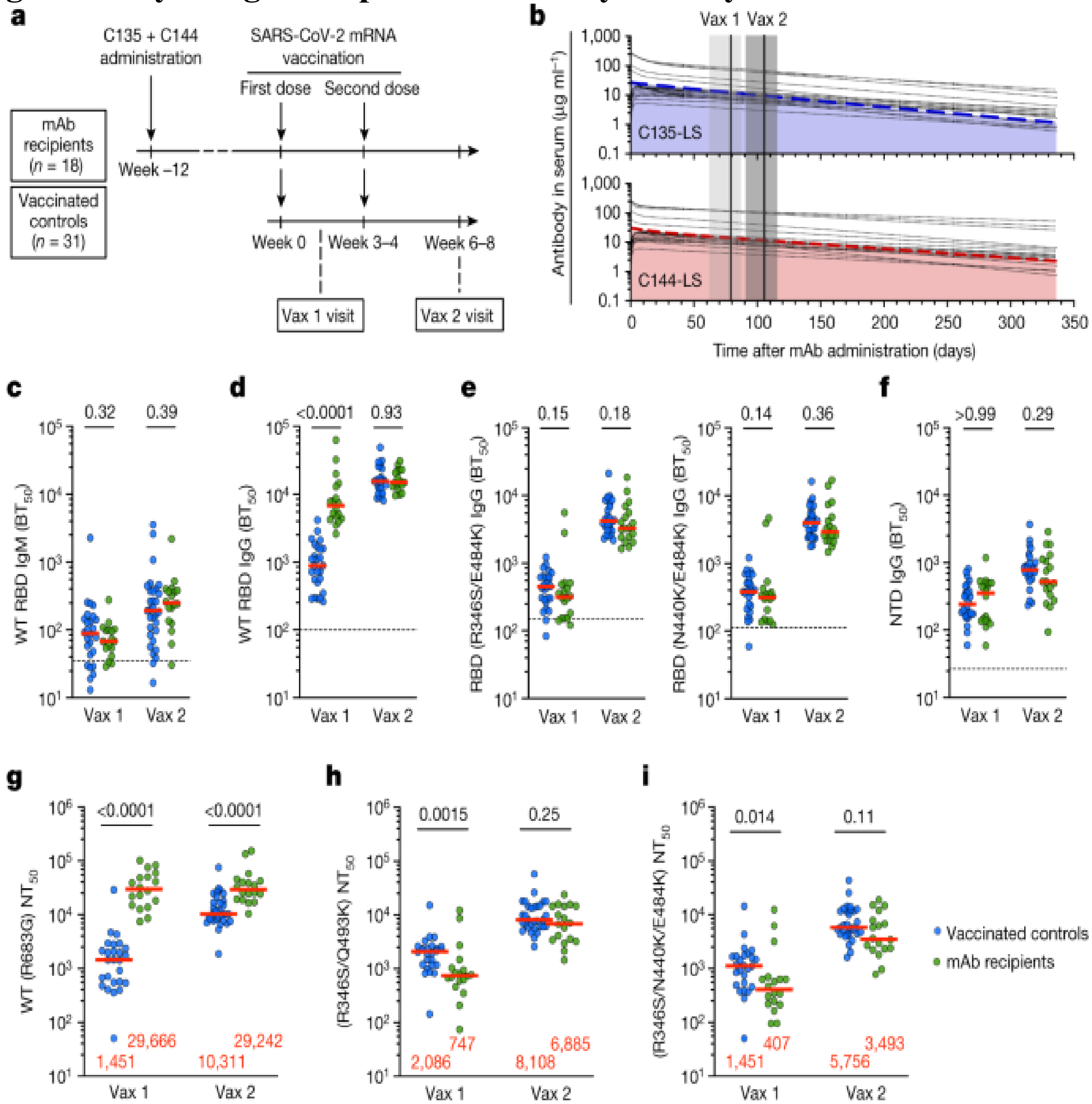
Feedback inhibition of humoral immunity by antibodies was first documented in 1909[1]. Subsequent studies showed that, depending on the context, antibodies can enhance or inhibit immune responses[2,3]. However, little is known about how pre-existing antibodies influence the development of memory B cells. Here we examined the memory B cell response in individuals who received two high-affinity anti-SARS-CoV-2 monoclonal antibodies and subsequently two doses of an mRNA vaccine[4,5,6,7,8]. We found that the recipients of the monoclonal antibodies produced antigen-binding and neutralizing titres that were only fractionally lower compared than in control individuals. However, the memory B cells of the individuals who received the monoclonal antibodies differed from those of control individuals in that they predominantly expressed low-affinity IgM antibodies that carried small numbers of somatic mutations and showed altered receptor binding domain (RBD) target specificity, consistent with epitope masking. Moreover, only 1 out of 77 anti-RBD memory antibodies tested neutralized the virus. The mechanism underlying these findings was examined in experiments in mice that showed that germinal centres formed in the presence of the same antibodies were dominated by low-affinity B cells. Our results indicate that pre-existing high-affinity antibodies bias germinal centre and memory B cell selection through two distinct mechanisms: (1) by lowering the activation threshold for B cells, thereby permitting abundant lower-affinity clones to participate in the immune response; and (2) through direct masking of their cognate epitopes. This may in part explain the shifting target profile of memory antibodies elicited by booster vaccinations[9].

Main

To examine how passive administration of monoclonal antibodies might influence subsequent humoral responses to vaccination in humans, we studied a group of 18 healthy volunteers who received a single dose of a combination of two long-acting monoclonal antibodies against SARS-CoV-2 and subsequently received two doses of a SARS-CoV-2 mRNA vaccine (Fig. 1a). The 2 antibodies—C144-LS and C135-LS—bind to class 2 and 3

epitopes on the RBD of the SARS-CoV-2 spike (S) protein with high affinity (dissociation constant (K_d) = 18 nM and K_d = 6 nM, respectively) and neutralize the virus with half-maximal inhibitory concentration (IC_{50}) values of 2.55 and 2.98 ng ml^{-1}, respectively[5,8].

Fig. 1: Study design and plasma antibody activity.



a, Schematic of the study design, with markers denoting weeks relative to the time of the first vaccine dose. mAb, monoclonal antibody. **b**, Serum levels of C135-LS (top, blue) and C144-LS (bottom, red) over time are shown. The thick coloured dashed lines indicate the median serum

concentrations among monoclonal antibody recipients (*n* = 18), and the thin dotted black lines represent individual participants. The two solid vertical lines indicate the median value and the grey shaded areas indicate the range of time from monoclonal antibody administration to vaccination. **c**–**f**, The half-maximal plasma binding titre (BT_{50}) to RBD after one (vax 1) and two doses (vax 2) of mRNA vaccination for monoclonal antibody recipients (*n* = 18, green) and controls (*n* = 26, blue). Each dot represents one individual. The dashed horizontal lines represent the median binding activity of pre-pandemic plasma samples from healthy individuals that were used as negative controls. **c**,**d**, IgM (**c**) and IgG (**d**) binding titres to WT RBD. **e**, IgG binding to R346S/E484K (left) and N440K/E484K RBDs (Extended Data Fig. 1). **f**, IgG binding to the NTD. **g**–**i**, Plasma half-maximal neutralizing titre (NT_{50}) values for monoclonal antibody recipients (*n* = 18, green) and controls (*n* = 26, blue) against HIV-1 pseudotyped with SARS-CoV-2 WT S (**g**), R346S/Q493K mutant S (**h**) and R346S/N440K/E484K mutant S (**i**) (Extended Data Fig. 2). The S protein in the pseudoviruses in **g**–**i** contained an R683G substitution. The red horizontal bars in **c**–**i** and the red numbers in **g**–**i** represent the median values. Statistical significance in **c**–**i** was determined using two-tailed Mann–Whitney *U*-tests comparing the differences between monoclonal antibody recipients and controls for each timepoint independently; *P* values are shown above the plots. All of the experiments were performed at least in duplicate.

Between 13 January and 3 March 2021, 23 SARS-CoV-2-naive individuals received C144-LS and C135-LS (*n* = 21) or placebo (*n* = 2) in a first-in-human phase 1 clinical trial at the Rockefeller University Hospital (ClinicalTrials.gov: NCT04700163). The antibodies were modified to extend their half-life by introducing the M428L and N343S mutations into their Fc domains[10] (LS). Individuals received a single dose of C144-LS and C135-LS IgG1 antibodies at a 1:1 ratio, starting with 100 mg of each subcutaneously (s.c.) in the lowest-dose group and up to 15 mg per kg intravenously (i.v.) in the highest-dose group. The participants were followed longitudinally to assess the safety and tolerability of the infused monoclonal antibodies and determine their pharmacokinetic properties.

A total of 18 out of the 21 phase 1 study participants who had received the monoclonal antibodies elected to receive SARS-CoV-2 mRNA vaccination

and volunteered to enrol in a parallel observational study assessing their immune responses to SARS-CoV-2 vaccination (Supplementary Tables 1 and 2). The first and second vaccine doses were administered a median of 82 (range, 42–110) and 103 (range, 70–131) days after antibody administration (Supplementary Tables 1 and 2). At the time of vaccination, the plasma levels of C144-LS and C135-LS were between 5 and 100 μg ml^{-1} depending on the dosing group (Fig. 1b). The estimated half-lives of C144-LS and C135-LS were 69–99 days and 73–95 days, respectively.

The 18 vaccinated antibody recipients were compared with a cohort of 31 randomly selected mRNA vaccine recipients with no previous history of infection[9,11] (Fig. 1a and Supplementary Table 1). Both of the groups were sampled 13–28 (median, 19) and 15–91 (median, 29) days after their first and second vaccine doses, respectively. The two cohorts were relatively matched for demographic characteristics and vaccine formulation (Supplementary Tables 1 and 2), and none of the individuals included in the study seroconverted to nucleocapsid (N) at any time during the observation period, suggesting that they remained infection naive.

Plasma antibody reactivity

Plasma IgM and IgG antibody-binding activity against RBD were measured with an enzyme-linked immunosorbent assay (ELISA) using Wuhan-Hu-1 (wild type (WT)) and mutant forms of RBD (R346S/E484K and N440K/E484K) that eliminate binding by C144 and C135 but not class 1 or 4, or some affinity-matured class 2 or 3 antibodies[9,12] (Extended Data Fig. 1a–f). When measured for WT RBD binding after one or two vaccine doses, the IgM titres of the monoclonal antibody recipients were not significantly different compared with the controls (Fig. 1c). By contrast, IgG anti-WT-RBD titres were significantly higher in monoclonal antibody recipients compared with in the controls after one vaccine dose, but equalized after the second dose (Fig. 1d; $P < 0.0001$ and $P = 0.93$, respectively). The initial difference was attributed to the infused monoclonal antibodies because, when the same samples were tested against either R346S/E484K or N440K/E484K mutant RBDs that interfere with C144 and C135 binding, plasma IgG antibody levels in the monoclonal antibody recipient samples were slightly, albeit not significantly, lower than the controls (Fig. 1e). By

contrast, the relative contribution of endogenous anti-RBD antibodies increased over time, as illustrated by a decrease in the correlation between the monoclonal antibody serum levels and plasma binding (Extended Data Fig. 1g,h).

To determine whether the pre-existing antibodies against the RBD interfered with humoral immunity to independent domains of the SARS-CoV-2 S protein, the same plasma samples were tested for binding to the N-terminal domain (NTD). IgG titres to the NTD were similar in the monoclonal antibody recipients and controls (Fig. 1f). We conclude that high circulating levels of C144-LS and C135-LS do not interfere with IgM anti-RBD antibody responses and have only a small effect on IgG responses. Thus, the infused antibodies do not clear the vaccine antigen or measurably interfere with its overall ability to produce an immune response[13].

Neutralization

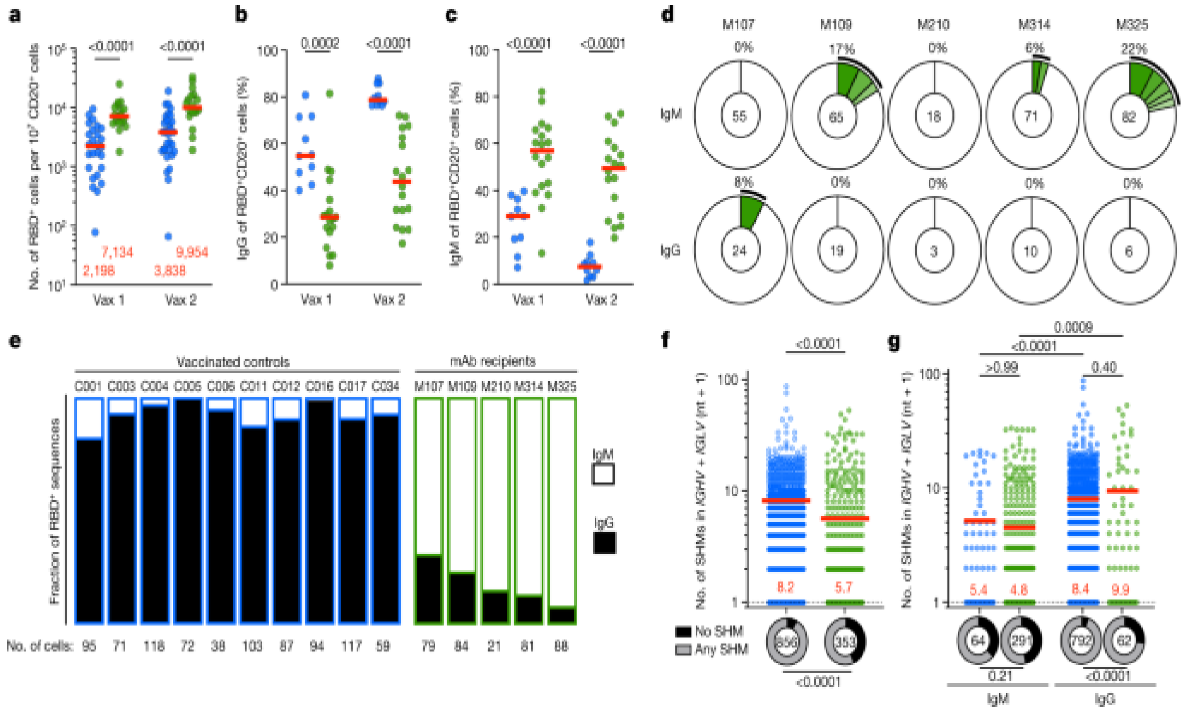
To assess plasma neutralizing activity, we used HIV-1 pseudotyped with WT or mutant S proteins carrying a mutation in the furin-cleavage site (R683G)[14]. As expected, on the basis of the amount of C144-LS and C135-LS in circulation, neutralizing titres against the WT were significantly higher in monoclonal antibody recipients compared with in the controls at both timepoints (Fig. 1g; $P < 0.0001$). To determine the contribution of the endogenous neutralizing response to epitopes outside of the C144-LS and C135-LS target sites, we used viruses pseudotyped with S proteins containing the R346S/Q493K and R346S/N440K/E484K mutations that abolish the neutralizing activity of the two infused monoclonal antibodies (Extended Data Fig. 2a–d). Despite the initial dominance of class 1–2 epitopes among neutralizing antibodies elicited by mRNA vaccination[6], the neutralizing titre of the control plasmas against the two mutant pseudoviruses were comparable to those against the WT (Fig. 1h,i), suggesting that a significant proportion of circulating endogenous neutralizing antibodies are unaffected by the R346S/Q493K and R346S/N440K/E484K mutations. After the first vaccine dose, monoclonal antibody recipients showed significantly lower neutralizing titres against the mutant pseudoviruses compared with the controls (2.7-fold and 3.5-fold for R346S/Q493K and R346S/N440K/E484K; Fig. 1h,i; $P = 0.0015$ and $P =$

0.014, respectively). Consistent with a recent report[13], neutralizing activity improved and was no longer significantly different compared with the controls after the second vaccine dose (Fig. 1h,i). Similarly, we saw no discernible difference in plasma neutralization of the antigenically divergent BA.4/5 variant, which showed equal levels of immune evasion to the plasma antibodies in both groups (Extended Data Fig. 2e). In conclusion, recipients of C144-LS and C135-LS had high initial levels of serum neutralizing activity due to the passively administered antibodies and they developed their own neutralizing antibodies that were not sensitive to RBD mutations in the C144/C135 target sites after mRNA vaccination.

Memory B cells

In addition to the plasma cells that produce circulating antibodies, vaccination also elicits memory B cells that contribute to protection after re-exposure to the pathogen. These two cell types are selected by different mechanisms and, as a result, the antibodies that they produce show differing levels of affinity to the immunogen[15,16,17]. Although the feedback effects of antibodies on humoral responses have been investigated extensively, beginning in 1909[1,2], little is known about their effects on the development of memory B cells. To investigate the effects of passive monoclonal antibody administration on B cell memory responses in humans, we used flow cytometry to enumerate and purify circulating memory B cells binding to phycoerythrin (PE)- and Alexa-Fluor-647 (AF647)-labelled RBDs[5] (Extended Data Fig. 3a–c). mRNA vaccination elicited robust RBD-specific memory B cell responses in monoclonal antibody recipients that were approximately fourfold and threefold higher than in the controls after the first and second vaccine doses, respectively (Fig. 2a; $P < 0.0001$ and $P < 0.0001$, respectively). Thus, administration of C144-LS and C135-LS increases the magnitude of the anti-RBD memory B cell response compared with the controls.

Fig. 2: Anti-SARS-CoV-2 RBD memory B cells from vaccinated monoclonal antibody recipients.



a–**c**, Flow cytometry enumeration and surface immunoglobulin expression of SARS-CoV-2 RBD-specific memory B cells after vax 1 and vax 2, isolated from monoclonal antibody recipients (green, $n = 18$) and control individuals[9,11] (blue). $n = 26$ for vax 1 and $n = 31$ for vax 2 (**a**), and $n = 10$ (**b** and **c**). Each dot represents one individual. The red horizontal bars (and the numbers in **a**) show the median values. **a**, The number of WT RBD-specific memory B cells per 10 million $CD20^+$ B cells (Extended Data Fig. 3a,b). **b**,**c**, The percentage of cells among WT RBD-binding $CD20^+$ B cells that express cell surface IgG (**b**) or IgM (**c**). **d**, The distribution of antibody sequences derived from cells isolated from five vaccinated monoclonal antibody recipients after vax 2 (Extended Data Fig. 4a–d). The numbers in the inner circles indicate the number of sequences analysed for the respective individual. The green coloured slices indicate clonally expanded cells (same *IGHV* and *IGLV* genes, with highly similar CDR3s) within an individual. Slice size is proportional to the number of clonally related sequences, with the fraction of clonally expanded sequences summarized as a percentage (black outline). The white areas of the pie chart indicate the proportion of sequences that were isolated only once. **e**, The fraction of cells containing IgG (black) versus IgM (white) transcripts per individual (see also Extended Data Fig. 4e and refs. [9,11]). **f**,**g**, Somatic hypermutation (SHM) shown as combined heavy- and light-chain variable region

nucleotide (nt) substitutions plus one (*IGVH* + *IGVL* + 1), with each dot representing one sequence from monoclonal antibody recipients (green) or controls (blue). The ring plots at the bottom show the fraction of sequences with no (*IGVH* + *IGVL* + 1 = 1) versus any (*IGVH* + *IGVL* + 1 > 1) somatic hypermutation, and the encircled numbers indicate the number of sequences analysed for all cells irrespective of isotype (**f**), and for IgM and IgG analysed independently (**g**). The red horizontal bars and numbers in **f** and **g** indicate the mean values. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests (**a**–**c** and **f**), Kruskal–Wallis tests with subsequent Dunn's correction for multiple comparisons (**g**) and two-sided Fisher's exact tests to compare fractions (**f** and **g**).

Human memory B cells represent a diverse pool of cells that can develop in germinal centres or through an extrafollicular germinal-centre-independent pathway[18,19,20]. Memory B cells expressing class-switched and highly somatically mutated antibodies are primarily of germinal centre origin[19,21]. IgM-expressing memory B cells, which express antibodies that carry only small numbers of mutations, typically develop through a germinal-centre-independent pathway[22,23,24,25]. In control individuals, IgG-expressing RBD-specific memory cells comprised the majority of the memory B cell pool at both timepoints assayed. Consistent with the overall increase in anti-RBD memory cells, the absolute number of IgG-expressing cells was fractionally increased in monoclonal antibody recipients (Extended Data Fig. 3d). However, their relative contribution was significantly reduced at both timepoints, making up only 30% and 45% of the RBD-specific cells in monoclonal antibody recipients after one and two doses, respectively (Fig. 2b; $P = 0.0002$ and $P < 0.0001$). Consistent with the relative decrease in IgG^+ memory B cells, more than half (57%) of the RBD-specific memory B cells from monoclonal antibody recipients were positive for cell-surface IgM after the first vaccine dose and this decreased only slightly to 49% after the second vaccine dose. By contrast, few such cells were found in the control group at that time (Fig. 2c and Extended Data Fig. 3e; all $P < 0.0001$). The skewed isotype ratio was correlated with the C144 serum concentration at the time of immunization (Extended Data Fig. 3f–i). We conclude that pre-existing high-affinity anti-RBD antibodies alter the immune response to SARS-CoV-2 mRNA vaccination to favour the development of IgM-expressing memory B cells.

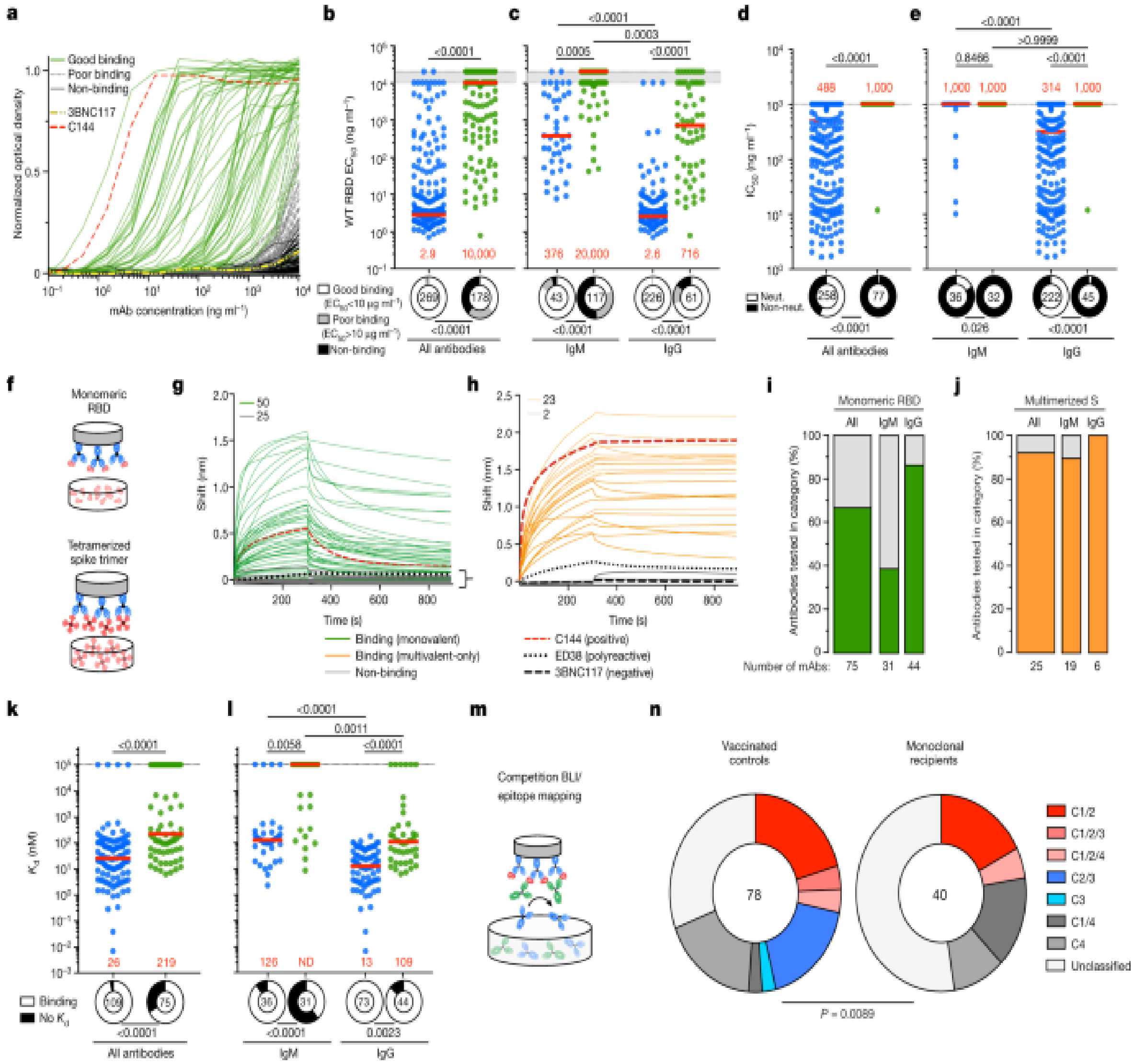
Memory B cell antibodies

To gain further insights into the effects of pre-existing antibodies on the human memory response to SARS-CoV-2 mRNA vaccination, we purified RBD-specific memory B cells from five representative monoclonal antibody recipients after the second vaccine dose (Extended Data Fig. 4a,b). A total of 353 and 856 paired antibody sequences from monoclonal antibody recipients and previously characterized controls were examined, respectively[11] (Fig. 2d, Extended Data Fig. 4c–e and Supplementary Table 3). IgM transcripts accounted for 70–94% of sequences recovered from monoclonal antibody recipients with an average of 9% belonging to expanded clones (Fig. 2d (top) and 2e). By contrast, IgG transcripts accounted for more than 90% of the immunoglobulin sequences isolated from the controls (Fig. 2e and Extended Data Fig. 4e). The relative IgM enrichment was more pronounced by the more sensitive PCR assay and may include cells that no longer express IgM on their surface. IgM memory cells originating from the extrafollicular non-germinal-centre pathway are generally less somatically mutated than IgG memory cells because they undergo fewer divisions[19,25]. Consistent with this idea, and the reversed ratio of IgM to IgG memory B cells in monoclonal antibody recipients, the antibodies obtained from these individuals showed significantly fewer somatic mutations compared with the controls (Fig. 2f; $P < 0.0001$). However, when comparing IgM or IgG cells independently, the average mutational burden was not significantly different between the monoclonal antibody recipients and the controls (Fig. 2g; $P > 0.99$ and $P = 0.40$ for IgM and IgG, respectively). Thus, IgM- and IgG-expressing B cells in vaccinated individuals who had received C144-LS and C135-LS carry normal numbers of somatic mutations, but the relative ratio of the two memory cell types is reversed, which accounts for the overall lower level of mutation in their memory compartment. Finally, in contrast to the controls, there was no enrichment for the VH3-53, VH1-69, VH1-46 and VH3-66 heavy chains, which often target class 1 and 2 epitopes. Instead, there was relative enrichment for the *VH3-9*, *VH5-51*, *VH4-39* and *VH1-8* genes (Extended Data Fig. 4f). The limited number of cells sequenced precludes definitive conclusions about the precise clonotype distribution in this population, but the relative change in *VH* gene use frequency implies that B cell recruitment into the memory compartment of monoclonal antibody recipients is altered. In summary, the data suggest that pre-existing

antibodies can alter the cellular and molecular composition of the RBD-specific MBC compartment that develops in response to mRNA vaccination.

To examine the binding and neutralizing activity of the memory antibodies elicited by mRNA vaccination in C144-LS and C135-LS recipients, we produced 178 representative monoclonal antibodies obtained from five individuals as IgGs and tested them for binding to the WT SARS-CoV-2 RBD using ELISA (Fig. 3a–c and Supplementary Table 4). In contrast to the controls, from whom over 95% of the memory antibodies bound strongly to the RBD, monoclonal antibodies isolated from volunteers who received C144-LS and C135-LS showed diverse levels of binding activity. Approximately one-quarter (24%) of the antibodies displayed relatively poor binding, with ELISA half-maximal effective concentrations (EC_{50}) that were only slightly above our limit of detection, and a little over one-third (38%) showed no detectable binding above the background (Fig. 3a,b). Accordingly, the median (EC_{50}) of antibodies isolated from monoclonal antibody recipients was significantly higher than in the controls (Fig. 3b; $P < 0.0001$). Notably, this difference remained significant when the monoclonal antibodies isolated from IgM and IgG memory cells were analysed independently (Fig 3c; $P = 0.0005$ and $P < 0.0001$, respectively).

Fig. 3: Anti-SARS-CoV-2 RBD memory antibodies from vaccinated monoclonal antibody recipients.



a–**c**, Monoclonal antibody binding to WT RBD. **a**, ELISA binding of monoclonal memory antibodies derived from monoclonal antibody recipients is shown. Each curve represents one antibody. The green curves show EC_{50} values of <10 μg ml^{-1}; the grey dashed lines show EC_{50} values of >10 μg ml^{-1}; the solid black lines are antibodies that were below or equal to the negative control anti-HIV1 antibody 3BNC117 (thick yellow dashed line). C144 (thick, red dashed line) was used as a positive control. **b**, EC_{50} values derived from **a** for monoclonal antibody recipients (green) and controls (blue) for all antibodies, irrespective of isotype. **c**, EC_{50} values as in **b**, but IgM and IgG were analysed independently. The grey shaded area between the horizontal dotted lines indicates antibodies with EC_{50} > 10 μg

ml^{-1} (poor binding) and non-binding antibodies, arbitrarily grouped at 10 and 20 μg ml^{-1}, respectively. The ring plots summarize the fraction of all antibodies tested for the respective groups (encircled number). **d**, IC_{50} values for all monoclonal antibodies isolated from vaccinated monoclonal antibody recipients (green) or control individuals (blue). The ring plots illustrate the fraction of non-neutralizing (non-neut.; IC_{50} > 1,000 ng ml^{-1}) antibodies (black slices) among all antibodies tested for the respective group (encircled number). **e**, IC_{50} values as described in **d**, but IgM and IgG antibodies were analysed independently. **f**–**l**, Monoclonal antibody binding to monomeric and multimerized antigen by BLI. **f**, Schematic of monomeric binding measurements in which IgG was immobilized onto the biosensor chip and subsequently exposed to monomeric RBD (top), and multimeric binding using 6P-stabilized WT SARS-CoV-2 S protein trimers that had been tetramerized using streptavidin (bottom). **g**, BLI traces obtained under monovalent conditions as shown in **f** (top). Each curve represents one antibody. The coloured solid lines denote binding above the background represented by polyreactive antibody ED38[35] (dotted black line) and anti-HIV-1 antibody 3BNC117 (dashed black line). The grey lines show non-binding antibodies. C144 (thick, red dashed line) was used as a positive control. **h**, BLI traces as described in **g** for antibodies that showed no measurable binding in **g** and were subsequently tested for binding under polyvalent conditions as illustrated in **f** (bottom). **i**, The percentage of binding antibodies under monovalent conditions for all antibodies and by isotype. The values below the bars indicate the number of antibodies tested. **j**, The percentage of binding antibodies as described in **i** for the antibodies shown in **h**. **k**, K_d values derived under monomeric binding conditions in **g** for monoclonal antibody recipients (green) and controls (blue) irrespective of isotype. The ring plots illustrate the fraction of antibodies tested for the respective group (encircled number) that measurably bound to monomeric RBD (binding, white) and those for which a K_d value could not be established (no K_d, black). **l**, K_d values as described in **k** were analysed independently for IgM and IgG. **m**, Schematic of the BLI competition experiment: (1) the capture antibody of known epitope specificity (class-reference antibody) was bound to the biosensor chip; (2) exposed to antigen; and (3) the antibody of interest was added to the chip. **n**, The distribution of the epitopes targeted. The number in the centre is the number of antibodies

tested. Slices coloured in shades of red and blue represent class 1, 2 and 3 or combined epitopes, and shades of grey represent class-4-containing epitopes or epitopes that could not be classified. For **b**–**e**, **k** and **l**, the red horizontal bars and numbers represent the median values. ND, not determined. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests (**b**, **d** and **k**), Kruskal–Wallis tests with subsequent Dunn's correction for multiple comparisons (**c**, **e** and **l**), two-sided Fisher's exact tests (**d**, **e**, **k** and **l**) and the two-sided χ^2 contingency statistic (**b**, **c** and **n**).

Memory antibodies obtained from C144-LS and C135-LS recipients that bound to WT SARS-CoV-2 RBD with EC_{50} values of <10 μg ml^{-1} were tested for neutralizing activity against viruses pseudotyped with WT S. Whereas almost two-thirds (63%) of the IgG and 17% of the IgM antibodies isolated from controls showed measurable neutralizing activity, only 1 out of 45 IgG-derived antibodies and none of the 32 IgM-derived antibodies obtained from C144-LS and C135-LS recipients neutralized SARS-CoV-2 (Fig. 3d,e). Thus, the antibodies isolated from the RBD-specific memory B cell compartment of vaccinated monoclonal antibody recipients show significantly less neutralizing activity compared with the antibodies isolated from the controls. In circulation, IgM antibodies are pentamers that, in addition to their superior ability to fix complement, show increased apparent affinities. To test whether the binding and neutralizing properties of IgM antibodies would be improved when expressed as pentamers, we re-expressed 17 IgM memory antibodies (15 from monoclonal antibody recipients and 2 from controls) as pentameric IgMs. Although the pentamers obtained from monoclonal antibody recipients showed significantly improved binding to RBD by ELISA (Extended Data Fig. 5a and Supplementary Table 4; $P = 0.0004$), they remained unable to neutralize (Extended Data Fig. 5b and Supplementary Table 4). By contrast, both control IgM antibodies showed improved neutralizing potency when expressed as pentamers (Extended Data Fig. 5b).

To examine the affinity of the antibodies, we performed biolayer interferometry experiments (BLI) in which monoclonal antibodies were immobilized onto the biosensor chip and exposed to WT RBD monomers[12] (Fig. 3f,g). In contrast to antibodies derived from the control individuals, for which 96% of the antibodies tested displayed measurable affinities, only

two-thirds (67%) of the antibodies derived from monoclonal antibody recipients did so (Fig. 3g,i,k; $P < 0.0001$). When all of the antibodies were considered together, the median affinity (affinity constant (K_d)) differed by nearly one order of magnitude between monoclonal antibody recipients and the controls (Fig. 3k; $P < 0.0001$). Moreover, this difference remained significant when IgM and IgG monoclonal antibodies were considered independently (Fig 3l; $P = 0.0058$ and $P < 0.0001$, respectively), indicating that the lower affinities observed in the memory compartment of monoclonal antibody recipients cannot solely be explained by the preponderance of IgM.

C144-LS and C135-LS have the potential to form immune complexes with the vaccine antigen in vivo and present it as a multimer that could increase the apparent affinity of a B cell for the multimerized antigen by avidity effects. To determine whether memory-cell-derived antibodies from monoclonal antibody recipients with no apparent affinity to monomeric antigen would show binding under higher valency conditions, we exposed the immobilized monoclonal antibodies to biotin–streptavidin tetramerized trimers of S (Fig. 3f,h,j). Of the 25 antibodies with no apparent monomeric binding tested, 23 (92%) bound to multimerized S (Fig. 3h,j). We conclude that most of the anti-RBD antibodies isolated from monoclonal antibody recipients that did not show detectable binding to RBD monomers bind to multimerized antigen. Thus, the absence of binding to monomeric antigen is a consequence of the relatively low affinity of the memory antibodies derived from monoclonal antibody recipients and not due to altered specificity.

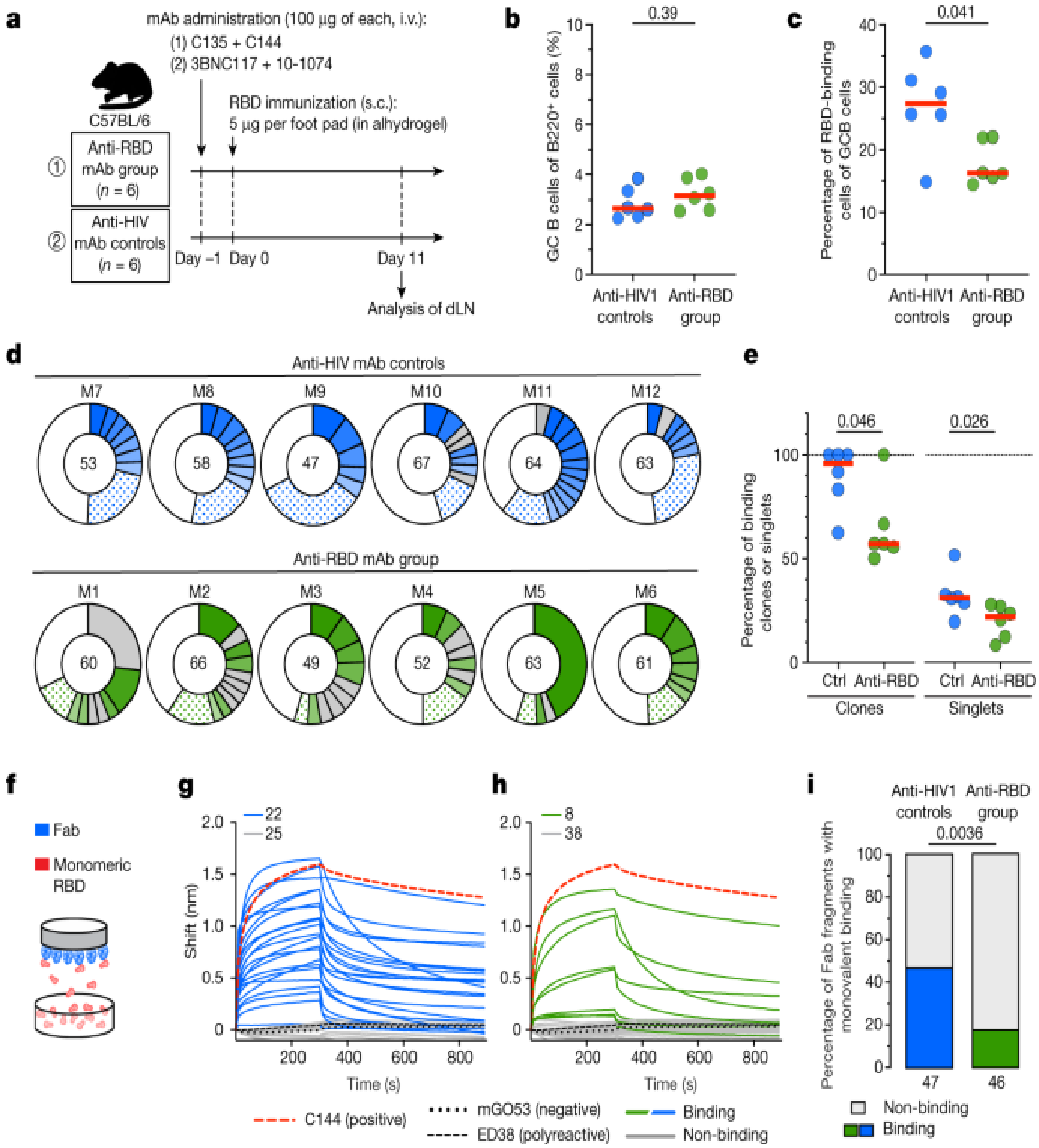
To examine the epitopes targeted by the vaccine-elicited anti-RBD antibodies produced by memory B cells of C144-LS and C135-LS recipients, we performed BLI experiments in which a preformed antibody–RBD complex comprising one of four structurally characterized antibodies[8,26,27] was exposed to a second monoclonal antibody targeting an unknown epitope (Fig. 3m and Extended Data Fig. 6). A total of 49% of the anti-RBD memory antibodies obtained from vaccinated controls target class 1, 2 or 3 epitopes or combinations thereof[9,11] (Fig. 3n and Extended Data Fig. 6a–f). By contrast, only 20% of the memory antibodies obtained from monoclonal antibody recipients targeted class 1 or 2 epitopes, and none were class 3 specific. Instead, we found that 78% of these antibodies targeted

either class-4-containing epitopes or epitopes that could not be classified using our method (Fig. 3n and Extended Data Fig. 6a–f). Thus, there was a significant shift in the distribution of epitopes targeted by memory antibodies isolated from monoclonal antibody recipients compared with the controls (Fig. 3n; $P = 0.0089$). In conclusion, C144-LS and C135-LS alter the development of memory B cells expressing anti-RBD antibodies and their epitope target preference.

Germinal centre B cell responses

To examine the mechanism by which pre-existing antibodies alter memory B cell selection, we immunized WT C57BL/6 mice that were preinfused with C135 and C144 or an irrelevant anti-HIV monoclonal antibody cocktail with recombinant SARS-CoV-2 RBD (Fig. 4a). Mice that were pretreated with control anti-HIV antibodies developed germinal centre responses in which an average of 27% of the B cells bound to RBD (Fig. 4b,c and Extended Data Fig. 7a–d). Although the germinal centre size was not altered in mice that received C135 and C144 (Fig. 4b and Extended Data Fig. 7c), the fraction of RBD-binding cells was significantly reduced (Fig. 4c and Extended Data Fig. 7b,d; $P = 0.041$).

Fig. 4: Germinal centre responses in mice pretreated with monoclonal antibodies.



a, Schematic of the experimental set-up. Pooled popliteal lymph nodes (dLN) were analysed using flow cytometry (Extended Data Fig. 7 and Methods). **b**,**c**, Enumeration of germinal centre B cells ($CD38^{-}FAS^{+}GL7^{+}$) as a fraction of all $B220^{+}$ B cells (**b**) and RBD-binding cells as a fraction of germinal centre B (GCB) cells (**c**). **d**,**e**, Antibody sequences from single germinal centre B cells. **d**, The distribution of antibody sequences. The encircled numbers indicate the number of sequences analysed per animal. Solid pie chart slices indicate clonally expanded sequences, with slices

coloured in shades of blue (controls) or green (anti-RBD monoclonal antibody group) indicating binding clones (Extended Data Fig. 7b and Supplementary Table 5). The grey slices denote non-binding clones. Sequences appearing only once are stippled (binding) or white (non-binding). **e**, The relative contribution of binding clones and singlets in **d**. **f**–**i**, Binding of monoclonal germinal centre B-cell-derived Fab fragments to monomeric RBD as determined using BLI. **f**, The BLI set-up. **g**,**h**, Traces of Fab fragments derived from controls (**g**) and anti-RBD-treated mice (**h**). Each curve represents one Fab. The coloured solid lines denote binding above the background represented by polyreactive antibody ED38[35] (dashed black line) and negative control antibody mGO53 (dotted black line). The grey lines show non-binding antibodies. C144 (thick, red-dashed line) was used as the positive control. **i**, The percentage of binding Fab fragments, with the total number of Fab fragments tested from the control ($n = 47$) and anti-RBD monoclonal antibody group ($n = 46$) denoted below. For **b**–**i**, control mice pretreated with irrelevant anti-HIV monoclonal antibodies (3BNC117 and 10-1074, $n = 6$) are shown in blue, and mice that received a combination of C135 and C144 ($n = 6$) are shown in green. For **b**, **c** and **e**, the coloured dots represent individual mice and the red horizontal lines indicate the median values. Statistical significance was determined using two-tailed Mann–Whitney *U*-tests (**b**, **c** and **e**) and two-sided Fisher's exact tests (**i**).

To examine the molecular nature of the memory antibodies produced in the presence of pre-existing high-affinity antibodies, we cloned and produced antibodies from germinal centre B cells. A total of 351 and 352 antibodies were obtained from the anti-RBD-treated and anti-HIV-treated control groups, respectively (Supplementary Table 5). B cells isolated from both groups showed similar levels of somatic mutation (Extended Data Fig. 7e) and clonal expansion (Fig. 4d and Extended Data Fig. 7f). However, in the C135 and C144 recipients, clonally expanded and unique B cells were dominated by cells that did not bind to the RBD, as determined using flow cytometry (Fig. 4d,e; $P = 0.046$ and $P = 0.026$, respectively).

A total of 106 monoclonal antibodies were expressed as fragment antigen-binding (Fab) fragments (Supplementary Table 6) and tested for binding to SARS-CoV-2 RBD by BLI (Fig. 4f). Under monomeric binding conditions, 46% (22 out of 47) of the Fab fragments derived from the control mice

bound to the RBD (Fig. 4g). By contrast, only 21% (8 out of 46) of the germinal centre B cell antibodies from the mice that were pretreated with C135 and C144 showed measurable but demonstrably lower affinity under these conditions (Fig. 4g–i; $P = 0.0036$ and Extended Data Fig. 7g). Thus, our mouse immunization experiments show that pre-existing high-affinity antibodies lower the affinity threshold for B cell participation in immune responses, and thereby provide a mechanistic explanation for the observation that the memory compartment in humans infused with C135 and C144 is dominated by low-affinity B cells.

Discussion

Our experiments show that pre-existing antibodies alter the development of memory B cells in response to SARS-CoV-2 mRNA vaccination in humans. Consistent with a recent report, C144-LS and C-135-LS did not significantly interfere with the development of circulating antibodies that bind to epitopes outside of the target sites of the two monoclonal antibodies[13]. Moreover, although we found that endogenous neutralizing responses were reduced in monoclonal antibody recipients after one dose, this difference was no longer significant after two doses of mRNA vaccination. By contrast, anti-RBD-specific memory B cell development was profoundly altered. Specifically, the affinity threshold for entry into the memory and germinal centre B cell compartment in humans and mice was lowered, and there was a change in the epitopes targeted in the presence of pre-existing antibodies.

Memory cells expressing IgG antibodies specific to class 1, 2 or 3 epitopes normally dominate the anti-RBD response after 2 doses of mRNA vaccination[4,5,6,7,8]. By contrast, we found that recipients of monoclonal antibodies developed increased numbers of IgM anti-RBD memory cells that express antibodies with altered epitope specificity consistent with epitope masking. This may also explain the shift in memory B cell specificity away from class 1 and 2 after the third dose of the SARS-CoV-2 mRNA vaccines that increases the breadth of the neutralizing response[9].

Beginning with experiments on anti-diphtheria toxin antibodies in the early part of the twentieth century, extensive work in experimental animals showed that passive transfer of polyclonal immune serum or monoclonal

antibodies can alter subsequent humoral immune responses in an epitope-specific manner[1,2]. More recently, epitope masking by pre-existing antibodies was shown to interfere with the epitope-specific plasmablast response to malaria vaccination[28,29]. In malaria vaccine trials, a third vaccine dose produced a smaller fraction of high-affinity antibody-producing plasmablasts than the second, and this was attributed to epitope masking. The masking effect of pre-existing high-affinity antibodies was confirmed in transgenic mice, showing that high-affinity antibodies block B cell entry into germinal centres in an epitope-specific manner[29,30,31]. However, these results were obtained in the context of a restricted transgenic B cell repertoire and the effect of pre-existing antibodies on the development of memory B cells in polyclonal responses was not examined[29]. The results of our clinical trial and mouse experiments agree with the observation that pre-existing antibodies block the development of epitope-specific B cells[32,33]. Our experiments extend previous observations because we also examined the specificity and antigen-binding affinity of the antibodies produced by B cells that do respond to antigen in the presence of pre-existing antibody. Notably, we found that pre-existing high-affinity antibodies lower the affinity threshold for B cell participation in the immune response. As a result, the anti-SARS-CoV2 antibodies expressed by memory B cells developing in humans who received monoclonal antibody infusions before vaccination are dominantly of low affinity. By contrast, low-affinity polyclonal antibodies emerging after a vaccine prime may enhance vaccine booster responses by mechanisms that have not yet been fully elucidated[2,3,31].

Memory B cells can develop through two different pathways[18,19]. Class-switched memory B cells that carry relatively higher-affinity antibodies with high numbers of somatic mutations develop in germinal centres. By contrast, IgM-expressing memory B cells that carry lower-affinity antibodies and only small numbers of mutations develop through a germinal-centre-independent pathway[18,20]. Our human data suggest that the passive transfer of C144-LS and C135-LS may favour the germinal-centre-independent pathway by creating immune complexes (reviewed in ref. [34]). Immune complexes trap and sequester antigen in lymphoid organs and thereby increase local antigen concentration. Furthermore, they contain multiple copies of the antigen in a form that increases apparent affinity by avidity effects, thereby enabling the

recruitment of B cells with very low-affinity receptors into the immune response[34]. Together, the increased antigen concentration and avidity effects combine to reduce selection stringency and help to explain the increase in low-affinity RBD-binding B cells found in human monoclonal antibody recipients, as this allows for the activation of greater numbers of abundant lower-affinity clones that would otherwise not be recruited to the immune response. Thus, the combination of epitope masking and lowered affinity thresholds diversify the ensuing B cell responses.

In conclusion, the composition of germinal centres and the development of memory in response to vaccination are influenced by pre-existing antibodies that can alter the antibody target profile, affinity, and isotype of the responding cells. Diversification of the antibody response through this mechanism may help to increase the breadth of vaccines like the SARS-CoV-2 vaccine but interfere with the development of breadth and potency in others, like HIV-1 or influenza, by diverting immunity away from broadly neutralizing to strain-specific epitopes and by allowing increasing numbers of low-affinity precursors expressing off-target antibodies to participate in the immune response.

Methods

Study participants

The participants in the monoclonal antibody recipient group were healthy SARS-CoV-2-naive volunteers who were enrolled in a first-in-human phase 1 study at the Rockefeller University Hospital in New York and received single doses of two anti-SARS-CoV-2 RBD monoclonal antibodies, C144-LS and C135-LS, between 13 January and 3 March 2021 (NCT04700163). The phase 1 clinical trial had a dose-escalation design and evaluated the safety and tolerability as well as the pharmacokinetics of the two monoclonal antibodies. The monoclonal antibody cocktail (1:1 ratio of C144-LS and C135-LS) was administered to 21 out of the 23 enrolled individuals ($n = 2$ receiving placebo), allowing for multiple interim safety analyses. The monoclonal antibodies were administered at 100 or 200 mg each s.c., or at 1, 5 or 15 mg per kg each i.v. (Supplementary Table 2). Eligible participants for the phase 1 study were healthy adults with no

history of SARS-CoV-2 infection or vaccination, or previous receipt of any SARS-CoV-2 therapeutics, including other monoclonal antibodies or plasma from convalescent individuals. Further details on the inclusion and exclusion criteria, study design and end points of the phase 1 study can be found at ClinicalTrials.gov (NCT04700163). Of the 21 individuals who received the monoclonal antibodies in the phase 1 study, 18 co-enrolled in a parallel observational study to assess their immune responses to subsequent SARS-CoV-2 mRNA vaccination. One individual chose to receive the Janssen (Ad26.COV2.S) vaccine, and another individual (a placebo recipient) displayed N titre changes before enrolment in the observational study that were compatible with a recent SARS-CoV-2 infection, making them ineligible for inclusion in this study. The remaining phase 1 study participants chose not to enrol in the parallel study of immune responses. The 18 monoclonal antibody recipients included in this observational study received either the Moderna (Spikevax, mRNA-1273) or Pfizer-BioNTech (Comirnaty, BNT162b2) mRNA vaccines against the WT (Wuhan-Hu-1) strain of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The participants in the vaccinated control group were healthy SARS-CoV-2-naive volunteers who had received two doses of one of the two currently approved SARS-CoV-2 mRNA vaccines, Moderna (Spikevax, mRNA-1273) or Pfizer-BioNTech (Comirnaty, BNT162b2). These control individuals had been recruited to the Rockefeller University Hospital for serial blood donations to longitudinally assess their immune responses to SARS-CoV-2 mRNA vaccination[9,11]. We previously reported the findings obtained from this group and refer to refs. [9,11] for further details on participant recruitment, inclusion and exclusion criteria, and demographic characteristics (Supplementary Tables 1 and 2). At each sample collection visit, the participants of either group presented to the Rockefeller University Hospital for blood sample collection and were asked to provide details of their vaccination regimen, possible side effects, comorbidities and possible COVID-19 history. Vaccinations were administered outside of the study, at the discretion of the individual and their health care provider consistent with existing guidelines and, as such, were not influenced by their participation in our study. Baseline and longitudinal plasma samples were tested for binding activity to the nucleocapsid protein (N; Sino Biological, 40588-V08B) of SARS-CoV-2. The absence of seroconversion towards N during the study interval was used to exclude SARS-CoV-2 infection, in addition to

participants' reported history. Clinical data collection and management were carried out using the software iRIS by iMedRIS (v.11.02). All of the participants provided written informed consent before participation in the study, which was conducted in accordance with good clinical practice. The study was performed in compliance with all of the relevant ethical regulations, and the clinical protocols (CGA-1015 and DRO-1006) for studies with human participants were approved by the Institutional Review Board of the Rockefeller University. Detailed participant characteristics are provided in Supplementary Tables 1 and 2 and refs. [9,11].

Blood sample processing and storage

Peripheral blood mononuclear cells obtained from samples collected at Rockefeller University were purified as previously reported by gradient centrifugation and stored in liquid nitrogen in the presence of fetal calf serum (FCS) and dimethylsulfoxide[5]. Heparinized plasma and serum samples were aliquoted and stored at −20 °C or less. Before the experiments, aliquots of plasma samples were heat-inactivated (at 56 °C for 1 h) and then stored at 4 °C.

Pharmacokinetics of C144-LS and C135-LS

To evaluate the pharmacokinetic properties of the passively administered antibodies C144-LS and C135-LS, their serum antibody levels were measured on the day of antibody administration (day 0) at 1, 3, 6, 9 and 12 h after infusion, and on days 1, 3, 7, 14, 21, 28, 56, 84, 126, 168, 252 and 336. C144-LS and C135-LS levels in the serum were measured using tandem mass spectrometry (MS/MS). In brief, analytes were isolated from serum samples through immunocapture using streptavidin beads and biotinylated RBD protein. The isolated proteins were denatured with dithiothreitol, alkylated with iodoacetamide, and digested with trypsin. The final extract was analysed using high-performance liquid chromatography with column-switching and MS/MS detection using positive-ion electrospray. A linear, 1/concentration^2-weighted, least-squares regression algorithm was used for quantification. Non-compartmental analysis was used to estimate the pharmacokinetic parameters from measured serum levels of C144-LS and C135-LS. Phoenix WinNonlin (v.8.2) was used for the non-compartmental

analysis. The actual sample time after the administration of each monoclonal antibody was used for the estimation of serum pharmacokinetic parameters instead of nominal time. Half-life estimates were similar between administration routes for both C144-LS and C135-LS, indicating a half-life of 2–3 months for both monoclonal antibodies by either administration route (C144-LS: 68.9 to 99.3 days for s.c. groups and 86.9 to 92.3 days for i.v. groups; C135-LS: 72.7 to 77.9 days for s.c. groups and 70.5 to 94.7 days for i.v. groups). Visualization of the pharmacokinetic data was performed in GraphPad Prism, using the three-phase decay model.

Mouse immunization

C57BL/6J mice were purchased from Jackson. All of the mice used were females aged 6–12 weeks at the start of the experiments. The mice were housed at a temperature of 22 ºC and a humidity of 30–70% under a 12 h–12 h light–dark cycle with ad libitum access to food and water. Footpad immunizations were performed using 25 µl of 1× PBS containing 5 µg of recombinant RBD and 8.5 µl of 2% Alhydrogel (Invivogen, 21645-51-2). 3BNC117/10-1074 and C135/C144 antibody cocktails were prepared with 100 µg of each antibody and delivered in 1× PBS through i.v. injection. All of the animal procedures and experiments were performed without blinding or randomization, and according to protocols approved by the Rockefeller University Institutional Animal Care and Use Committee (IACUC).

ELISAs

ELISA assays[36,37] to evaluate antibody binding to SARS-CoV-2 Wuhan-Hu-1 RBD, NTD or S were performed by coating high-binding 96-half-well plates (Corning 3690) with 50 µl per well of a 1 µg ml^{-1} protein solution in phosphate-buffered saline (PBS) overnight at 4 °C. The plates were washed six times with washing buffer (1× PBS with 0.05% Tween-20 (Sigma-Aldrich)) and incubated with 170 µl per well blocking buffer (1× PBS with 2% BSA and 0.05% Tween-20 (Sigma-Aldrich)) for 1 h at room temperature. Immediately after blocking, monoclonal antibodies or plasma samples were added in PBS and incubated for 1 h at room temperature. Plasma samples were assayed at a 1:66 starting dilution and serially diluted by either threefold or fourfold. Monoclonal antibodies were tested at a

starting concentration of 10 μg ml^{-1} and 11 additional threefold serial dilutions. The plates were washed six times with washing buffer and then incubated with anti-human IgG or IgM secondary antibody conjugated to horseradish peroxidase (HRP) (Jackson Immuno Research, 109-036-088 and 109-035-129) in blocking buffer at a 1:5,000 dilution (IgM and IgG). The plates were developed by addition of the HRP substrate, 3,3′,5,5′-tetramethylbenzidine (Thermo Fisher Scientific) for 6 min. The developing reaction was stopped by adding 50 μl of 1 M H_2SO_4 and absorbance was measured at 450 nm using an ELISA microplate reader (FluoStar Omega, BMG Labtech) with Omega and Omega MARS software for analysis. Normalizer control samples were included on each plate. For plasma samples and monoclonal antibodies, half-maximal binding titres (BT_{50}) and half-maximal effective concentrations (EC_{50}), respectively, were calculated using a four-parameter nonlinear regression model (GraphPad Prism v.9.3) with the following settings: [agonist] vs. response -- variable slope (four parameters); bottom = 0; Hillslope > 0; top = experiment-specific upper plateau of the normalizer control antibody or plasma sample reaching saturation for at least 3 consecutive dilution steps. The curve fit was constrained to an upper limit that corresponds to the maximal optical density achieved by the known normalizer control to limit interplate/interexperiment variability (batch effects). Pentameric IgM BT_{50} values were established using previously measured IgG antibodies as normalizer controls. Pre-pandemic plasma samples from healthy donors and isotype control monoclonal antibodies were used as negative controls as indicated and were used for validation[5]. All of the reported EC_{50} and BT_{50} values are the average of at least two independent experiments.

Proteins

The mammalian expression vector encoding the RBD of SARS-CoV-2 (GenBank MN985325.1; S protein residues 319–539) was previously described[26]. Plasmids encoding the R346S/E484K and N440K/E484K substitutions were generated using a site-directed mutagenesis kit according to the manufacturer’s instructions (New England Biolabs (NEB), E0554S). All of the constructs were confirmed by Sanger sequencing and were used to express soluble proteins by transiently transfecting Expi293F cells

(Gibco/Thermo Fisher Scientific, A14527)). Supernatants were collected after 4 days, and RBD proteins were purified by nickel affinity chromatography. S 6P proteins were purified by nickel affinity followed by size-exclusion chromatography. The peak fractions from size-exclusion chromatography were identified using native gel electrophoresis, and the peak fractions corresponding to monomeric RBDs or S trimers were pooled and stored at 4 °C.

SARS-CoV-2 pseudotyped reporter virus

A plasmid expressing SARS-CoV-2 S in the context of pSARS-CoV-2-$S_{\Delta19}$ (Wuhan-Hu-1) has been described previously[6], and a derivative of pSARS-CoV-2-$S_{\Delta19}$ with a disrupted furin-cleavage site was generated by introducing the R683G substitution[6]. Disruption of the furin-cleavage site results in increased particle infectivity. A plasmid expressing SARS-CoV-2 Omicron BA.4/5 S carrying the R683G substitution has been described previously[38]. Two plasmids containing C135/C144 antibody escape mutations were constructed on the basis of pSARS-CoV-2-$S_{\Delta19}$ R683G by overlap extension PCR-mediated mutagenesis and Gibson assembly. Specifically, the substitutions introduced were as follows: R346S/Q493K and R346S/N440K/E484K. Importantly, as those substitutions were incorporated into the pSARS-CoV-2-$S_{\Delta19}$ R683G background, neutralizing activity against all mutant and variant pseudoviruses were compared to a WT SARS-CoV-2 S sequence (GenBank: NC_045512) also carrying R683G (pSARS-CoV-2-$S_{\Delta19}$ R683G) where appropriate, as indicated. SARS-CoV-2 pseudotyped particles were generated as previously described[5,14]. In brief, HEK293T cells were transfected with pNL4-3ΔEnv-nanoluc and pSARS-CoV-2-$S_{\Delta19}$, and particles were collected 48 h after transfection, filtered and stored at −80 °C.

Pseudotyped virus neutralization assay

Fivefold serially diluted pre-pandemic negative control plasma from healthy donors (technical negative controls, data not shown), plasma from vaccinated monoclonal antibody recipients and mRNA vaccinated controls, or monoclonal antibodies were incubated with SARS-CoV-2 pseudotyped

virus for 1 h at 37 °C. The mixture was subsequently incubated with HEK293T-ACE2 cells[5] (for all monoclonal antibody WT neutralization assays) or HT1080Ace2 cl14 cells[6] (for all plasma neutralization assays) for 48 h, after which cells were washed with PBS and lysed with Luciferase Cell Culture Lysis 5× reagent (Promega). Nanoluc luciferase activity in the lysates was measured using the Nano-Glo Luciferase Assay System (Promega) with the ClarioStar Multimode reader (BMG, v.5.70.R3). The relative luminescence units were normalized to those derived from cells infected with SARS-CoV-2 pseudotyped virus in the absence of plasma or monoclonal antibodies. The half-maximal neutralization titre for plasma (NT_{50}) or half-maximal and 90% inhibitory concentrations for monoclonal antibodies (IC_{50} and IC_{90}) were determined using four-parameter nonlinear regression (least-squares regression method without weighting; constraints: top=1, bottom=0; in GraphPad Prism).

Biotinylation of viral protein for use in flow cytometry and BLI

Purified and Avi-tagged SARS-CoV-2 Wuhan-Hu-1 RBD or S were biotinylated using the Biotin-Protein Ligase-BIRA kit according to manufacturer's instructions (Avidity) as described previously[5]. Ovalbumin (Sigma-Aldrich, A5503-1G) was biotinylated using the EZ-Link Sulfo-NHS-LC-Biotinylation kit according to the manufacturer's instructions (Thermo Fisher Scientific). Biotinylated ovalbumin was conjugated to streptavidin-BV711 for human single-cell sorts (BD Biosciences, 563262) or to streptavidin-BB515 for phenotyping (BD Biosciences, 564453). For all human and mouse experiments, RBD was conjugated to streptavidin-PE (BD Biosciences, 554061) and streptavidin-AF647 (BioLegend, 405237)[5,9].

Human flow cytometry and single-cell sorting

Single-cell sorting by flow cytometry was described previously[5]. In brief, peripheral blood mononuclear cells were enriched for B cells by negative selection using a pan-B-cell isolation kit according to the manufacturer's instructions (Miltenyi Biotec, 130-101-638). Before staining, the enriched B cells were incubated with an FcR-blocking antibody (BD, 564220) at a 1:200 dilution in fluorescence-activated cell sorting (FACS) buffer (1× PBS,

2% FCS, 1 mM ethylenediaminetetraacetic acid (EDTA)) for 20 min on ice. Subsequently, cells were incubated in FACS buffer with the following anti-human antibodies (all at 1:200 dilution): anti-CD20-PECy7 (BD Biosciences, 335793), anti-CD3-APC-eFluro 780 (Invitrogen, 47-0037-41), anti-CD8-APC-eFluor 780 (Invitrogen, 47-0086-42), anti-CD16-APC-eFluor 780 (Invitrogen, 47-0168-41), anti-CD14-APC-eFluor 780 (Invitrogen, 47-0149-42) and Zombie NIR (BioLegend, 423105), and fluorophore-labelled RBD and ovalbumin (Ova) for 30 min on ice. AccuCheck Counting Beads (Life Technologies, PCB100) were added as indicated to each sample according to manufacturer's instructions. Single CD3^{-}CD8^{-}CD14^{-}CD16^{-}CD20^{+}Ova^{-}RBD–PE^{+}RBD−AF647^{+} B cells were sorted into individual wells of 96-well plates containing 4 µl of lysis buffer (0.5× PBS, 10 mM dithiothreitol, 3,000 U ml^{-1} RNasin Ribonuclease Inhibitors (Promega, N2615) per well using the FACS Aria III system and FACSDiva software (Becton Dickinson) for acquisition and FlowJo for analysis. The sorted cells were frozen on dry ice, and then stored at −80 °C for subsequent RNA reverse transcription. For B cell phenotype analysis, in addition to above antibodies, B cells were also stained with following anti-human antibodies (all at 1:200 dilution): anti-CD19-BV605 (BioLegend, 302244), anti-IgG-PECF594 (BD, 562538), anti-IgM-AF700 (BioLegend, 314538) and anti-CD38-BV421 (BioLegend, 303526).

Mouse flow cytometry and single-cell sorting

Popliteal lymph nodes from mice were collected in FACS buffer (1× PBS, 2% FBS, 2 mM EDTA), mechanically disrupted and filtered through a 35 µM strainer (Corning, 352235). Cells then underwent iterative rounds of staining each for 20 min on ice (all of the antibodies were diluted at 1:200 unless stated otherwise): (1) anti-mouse CD16/32 (Mouse BD FC Block, BD Biosciences, 553142); (2) fluorophore-conjugated RBD (see above); (3) anti-T and anti-B cell activation antigen-FITC (BD Biosciences, 553666), anti-CD38-PB (BioLegend, 102720; 1:100 dilution), anti-CD45R/B220-BV605 (BioLegend, 103244), anti-CD4-APC-eFluor780 (Invitrogen, 47-0042-82), anti-CD8a-APC-eFluor780 (Invitrogen, 47-0081-82), anti-NK1.1-APC-eFluor780 (Invitrogen, 47-5941-82), anti-F4/80-APC-eFluor780 (Invitrogen, 47-4801-82), anti-CD95-PE-Cy7 (BD Biosciences, 557653) and Zombie NIR (BioLegend, 423105, 1:1,000). AccuCheck Counting Beads

(Life Technologies, PCB100) were added to samples according to the manufacturer's instructions. Single $CD4^{-}CD8a^{-}NK1.1^{-}F4/80^{-}B220^{+}CD38^{-}GL7^{+}CD95^{+}$ cells were sorted on the BD FACSymphony S6 system into 96-well plates containing 1% 2-β-mercaptoethanol (Sigma-Aldrich) in TCL Buffer (Qiagen, 1031576) and subsequently frozen on dry ice and stored at −80 °C for RNA reverse transcription.

Antibody sequencing, cloning and expression

Human antibodies were identified and sequenced as described previously[5,39]. In brief, RNA from single cells was reverse-transcribed (SuperScript III Reverse Transcriptase, Invitrogen, 18080-044) and the cDNA was stored at −20 °C or used for subsequent amplification of the variable *IGH*, *IGL* and *IGK* genes by nested PCR and Sanger sequencing. Sequence analysis was performed using MacVector (v.17.5.4) and Geneious Prime (v.2020.1.2 and v.2022.1.1). Amplicons from the first PCR reaction were used as templates for sequence- and ligation-independent cloning into antibody expression vectors. Mouse monoclonal antibodies were sequenced and cloned from single FACS-sorted B cells as previously described[40,41] with the following modifications. In brief, RNA from single cells in 96-well plates was purified using magnetic beads (RNAClean XP, Beckman Coulter, A63987). Single-cell RNA was eluted from magnetic beads with 11 µl of solution containing 14.4 ng μl^{-1} of random primers (Invitrogen, 48190011), 0.5% of IGEPAL CA-630 (10% in distilled H–O, Sigma-Aldrich, I8896-50ML) and 0.6 U μl^{-1} of RNase inhibitor (Promega, N2615) in nuclease-free water (Qiagen, 129115), followed by incubation at 65 °C for 3 min. cDNA was synthesized by reverse transcription (SuperScript III Reverse Transcriptase 10,000 U, Invitrogen, 18080-044) and stored at −20 °C after addition of 10 µl nuclease-free water. Subsequent amplification of the variable *IGH* and *IGK* antibody genes was achieved by nested PCR using HotstarTaq DNA polymerase (Qiagen, 203209), using previously published primers[41] and the following thermocycler conditions for annealing (see details below), elongation (all 72 ºC/55 s) and number of cycles: PCR1 (*IGG*, *IGM* and *IGK*): 46 °C/30 s/50 cycles; PCR2 (*IGG* and *IGM*): 55 °C/30 s/50 cycles; PCR2 (*IGK*): 46 °C/30 s/50 cycles. After purification, heavy chain and light chain PCR products were Sanger sequenced, and

subsequently analysed using MacVector and Geneious Prime (v.2022.1.1), as well as the bioinformatics pipeline detailed below. Mouse Ig sequences were ordered as eBlocks (IDT) with short homologies for sequence and ligation independent cloning and cloned into human IGHG1 Fab and IGLK expression vectors as previously described[41]. All plasmid sequences were verified by Sanger sequencing, and all recombinant monoclonal antibodies (human memory B cell derived full-length IgG and His_6-tagged mouse-derived human IgG1 Fab fragments) were thereafter produced and purified as previously described[5,41]. To produce pentameric IgMs, cloning from PCR products was performed by sequencing and ligation-independent cloning as described above, except that the *IGH* variable gene was cloned into an Igμ expression vector (InvivoGen, pfusess-hchm3). Pentameric IgMs were then expressed through transient transfection of HEK293-6E cells with vectors encoding the appropriate J chain, light chain and heavy chain at a ratio of 1:1.5:1.5. Secreted IgMs were collected from cell supernatants after 6 days and purified using the POROS CaptureSelect IgM Affinity Matrix kit (Thermo Fisher Scientific, 1952890500). Affinity-purified IgMs were further purified by size-exclusion chromatography. Peak fractions of pentameric IgMs were analysed using SDS–PAGE, pooled and buffer-exchanged into PBS using an Amicon Ultra 100 kDa (Millipore) centrifugal filter unit.

BLI analysis

BLI assays were performed as previously described[5,15,41] with minor modifications as below. In brief, we used the ForteBio Octet Red instrument (ForteBio Data Acquisition software v.11.1.3.25) at 30 °C with shaking at 1,000 r.p.m. Monomeric affinities of anti-SARS-CoV-2 RBD IgG and Fab fragments to RBD were derived by subtracting the signal obtained from traces performed with the same IgG/Fab in the absence of WT RBD. Kinetic analysis using protein A biosensor (ForteBio, 18-5010) for human IgGs was performed as follows: (1) baseline: immersion for 60 s in buffer (1× Octet Kinetic buffer, Sartorius, 18-1105); (2) loading: immersion for 200 s in a solution with IgGs at 10 μg ml^{-1}; (3) baseline: immersion for 200 s in buffer; (4) association: immersion for 300 s in solution with WT RBD at three different concentrations ranging from 200 to 5 μg ml^{-1}; (5) dissociation: immersion for 600 s in buffer. Curve fitting was performed using a fast 1:1

binding model and the data analysis software from ForteBio. Mean equilibrium dissociation constants (K_d) were determined by averaging all binding curves that matched the theoretical fit with an R^2 value of ≥0.8. To establish binding of low-affinity antibodies to multimerized antigen, 6P-stabilized and biotinylated S trimers were incubated with recombinant streptavidin (ACROBiosystems, STN-N5116) for 30 min at room temperature, resulting in up to 12 RBD-binding moieties per molecule and assayed on the Octet Red instrument (ForteBio) as described above, with the following modification: association step (4) was performed with the S-multimer at 430 μg ml^{-1}. Epitope mapping assays were performed with protein A biosensor (ForteBio, 18-5010), according to the manufacturer's protocol 'classical sandwich assay' as follows: (1) sensor check: sensors were immersed for 30 s in buffer alone (buffer ForteBio, 18-1105); (2) capture first antibody: sensors were immersed for 10 min with antibody 1 at 10 μg ml^{-1}; (3) baseline: sensors were immersed 30 s in buffer alone; (4) blocking: sensors were immersed for 5 min with IgG isotype control (3BNC117) at 20 μg ml^{-1}; (5) baseline: sensors were immersed for 30 s in buffer alone; (6) antigen association: sensors were immersed for 5 min with RBD at 20 μg ml^{-1}; (7) baseline: sensors were immersed for 30 s in buffer alone. (8) Association antibody b2: sensors were immersed for 5 min with antibody 2 at 10 μg ml^{-1}. Affinity testing of mouse germinal centre B-cell-derived human IgG1 Fab fragments was performed using the same steps and Octet Red instrument (ForteBio) settings as for the human memory-derived full-length IgG antibodies (see above) with the following modifications: monoclonal Fab fragments at 50 μg ml^{-1} were captured on FAB2G biosensors (Sartorius 18-5125) for step 2; monovalent RBD was added at concentrations ranging from 30 to 120 μg ml^{-1} in step 4. Binding Fab fragments with measurable affinities were defined as having biologically plausible association and dissociation kinetics (that is, being able to associate to saturation on the biosensor in step 2 and showing a discernible association and dissociation of antigen after steps (4) and (5)) as well as a computed K_d value that matched the theoretical fit with an R^2 value of ≥0.75. Affinities of Fab fragments that did not get captured on the biosensor to saturation at the concentration tested could not be resolved and were excluded from further analysis (marked as N/D in Supplementary Table 6).

In all cases, curve fitting was performed using the ForteBio Octet Data analysis software (ForteBio Data Analysis HT v.11.1.3.50).

Computational analyses of antibody sequences

Antibody sequences were trimmed based on quality and annotated using Igblastn v.1.14 with the IMGT domain delineation system. Annotation was performed systematically using Change-O toolkit (v.0.4.540)[42]. The clonality of the heavy and light chain was determined using DefineClones.py implemented by Change-O (v.0.4.5)[42]. The script calculates the Hamming distance between each sequence in the dataset and its nearest neighbour. Distances are subsequently normalized to account for differences in junction sequence length, and clonality is determined based on a cut-off threshold of 0.15. Heavy and light chains derived from the same cell were subsequently paired, and clonotypes were assigned on the basis of their *V* and *J* genes using custom R and Perl scripts. All scripts and the data used to process antibody sequences are publicly available at GitHub (https://github.com/stratust/igpipeline/tree/igpipeline2_timepoint_v2). The frequency distribution of human *V* genes in anti-SARS-CoV-2 antibodies from this study (Extended Data Fig. 4f–h) was compared to 131,284,220 IgH and IgL sequences generated in ref. [43] and downloaded from cAb-Rep[44]—a database of human shared BCR clonotypes available online (https://cab-rep.c2b2.columbia.edu/). On the basis of the 108 distinct *V* genes that make up the 417 analysed sequences from the Ig repertoire of the individuals described in this study (353 sequences isolated form 5 monoclonal antibody recipients and 65 IgM sequences isolated from 9 vaccinated control individuals (for IgG sequences isolated from controls, see refs. [9,11]), we selected the IgH and IgL sequences from the database that are partially coded by the same *V* genes and counted them according to the constant region. The frequencies shown in Extended Data Fig. 4f–h are relative to the source and isotype analysed. We used two-sided binomial tests to check whether the number of sequences belonging to a specific *IGHV* or *IGLV* gene in the repertoire was different according to the frequency of the same *IGV* gene in the database. Adjusted *P* values were calculated using correction for false-discovery rate. Significant differences are denoted by asterisks. Nucleotide somatic mutations and complementarity-determining region (CDR3) length were determined using custom R and Perl scripts. For

quantification of somatic mutations, the *IGHV* and *IGLV* nucleotide sequences were aligned against their closest germlines using Igblastn and the number of differences was considered to be the number of nucleotide mutations.

Data presentation

Figures were arranged in Adobe Illustrator 2022.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are provided in Supplementary Tables 1–6. The raw sequencing data and computer scripts associated with Figs. 2 and 4 and Extended Data Figs. 3 and 7 have been deposited at GitHub (https://github.com/stratust/igpipeline/tree/igpipeline2_timepoint_v2). This study also uses data from 'A public database of memory and naive B-cell receptor sequences' (https://doi.org/10.5061/dryad.35ks2), the Protein Data Bank (6VYB and 6NB6), cAb-Rep (https://cab-rep.c2b2.columbia.edu/), the Sequence Read Archive (SRP010970) and from 'High frequency of shared clonotypes in human B cell receptor repertoires' (https://doi.org/10.1038/s41586-019-0934-8).

Code availability

Computer code to process the antibody sequences is available at GitHub (https://github.com/stratust/igpipeline/tree/igpipeline2_timepoint_v2).

References

1. Smith, T. Active immunity produced by so called balanced or neutral mixtures of diphtheria toxin and antitoxin. *J. Exp. Med.* **11**, 241–256

(1909).

2. Heyman, B. Regulation of antibody responses via antibodies, complement, and Fc receptors. *Annu. Rev. Immunol.* **18**, 709–737 (2000).

3. Bournazos, S. & Ravetch, J. V. Fcgamma receptor function and the design of vaccination strategies. *Immunity* **47**, 224–233 (2017).

4. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).

5. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

6. Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).

7. Yuan, M. et al. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **373**, eabh1139 (2021).

8. Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

9. Muecksch, F. et al. Increased memory B cell potency and breadth after a SARS-CoV-2 mRNA boost. *Nature* **607**, 128–134 (2022).

10. Zalevsky, J. et al. Enhanced antibody half-life improves in vivo activity. *Nat. Biotechnol.* **28**, 157–159 (2010).

11. Cho, A. et al. Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination. *Nature* **600**, 517–522 (2021).

12. Wang, Z. et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021).

13. Benschop, R. J. et al. The anti-SARS-CoV-2 monoclonal antibody bamlanivimab minimally affects the endogenous immune response to

COVID-19 vaccination. *Sci. Transl. Med.* **14**, eabn3041 (2022).

14. Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, e20201181 (2020).

15. Viant, C. et al. Antibody affinity shapes the choice between memory and germinal center B cell fates. *Cell* **183**, 1298–1311 (2020).

16. Ise, W. et al. T follicular helper cell-germinal center B cell interaction strength regulates entry into plasma cell or recycling germinal center cell fate. *Immunity* **48**, 702–715 (2018).

17. Phan, T. G. et al. High affinity germinal center B cells are actively selected into the plasma cell compartment. *J. Exp. Med.* **203**, 2419–2424 (2006).

18. Kurosaki, T., Kometani, K. & Ise, W. Memory B cells. *Nat. Rev. Immunol.* **15**, 149–159 (2015).

19. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **40**, 413–442 (2022).

20. Weisel, F. & Shlomchik, M. Memory B cells of mice and humans. *Annu. Rev. Immunol.* **35**, 255–284 (2015).

21. Klein, U., Rajewsky, K. & Küppers, R. Human immunoglobulin (Ig)$M^{+}IgD^{+}$ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J. Exp. Med.* **188**, 1679–1689 (1998).

22. Klein, U., Küppers, R. & Rajewsky, K. Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood* **89**, 1288–1298 (1997).

23. Taylor, J. J., Pape, K. A. & Jenkins, M. K. A germinal center-independent pathway generates unswitched memory B cells early in the primary response. *J. Exp. Med.* **209**, 597–606 (2012).

24. Viant, C. et al. Germinal center-dependent and -independent memory B cells produced throughout the immune response. *J. Exp. Med.* **218**, e20202489 (2021).

25. Weisel, F. J., Zuccarino-Catania, G. V., Chikina, M. & Shlomchik, M. J. A temporal switch in the germinal center determines differential output of memory B and plasma cells. *Immunity* **44**, 116–130 (2016).

26. Barnes, C. O. et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* **182**, 828–842 (2020).

27. Jette, C. A. et al. Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *Cell Rep.* **36**, 109760 (2021).

28. Kurtovic, L., Boyle, M. J. & Beeson, J. G. Epitope masking may limit antibody boosting to malaria vaccines. *Immunol. Cell Biol.* **99**, 126–129 (2021).

29. McNamara, H. A. et al. Antibody feedback limits the expansion of B cell responses to malaria vaccination but drives diversification of the humoral response. *Cell Host Microbe* **28**, 572–585 (2020).

30. Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. & Jenkins, M. K. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* **331**, 1203–1207 (2011).

31. Tas, J. M. J. et al. Antibodies from primary humoral responses modulate recruitment of naive B cells during secondary responses. *Immunity* **55**, 1856–1871 (2022).

32. Meyer-Hermann, M. Injection of antibodies against immunodominant epitopes tunes germinal centers to generate broadly neutralizing antibodies. *Cell Rep.* **29**, 1066–1073 (2019).

33. Zarnitsyna, V. I., Lavine, J., Ellebedy, A., Ahmed, R. & Antia, R. Multi-epitope models explain how pre-existing antibodies affect the

generation of broadly protective responses to influenza. *PLoS Pathog.* **12**, e1005692 (2016).

34. Heesters, B. A., Myers, R. C. & Carroll, M. C. Follicular dendritic cells: dynamic antigen libraries. *Nat. Rev. Immunol.* **14**, 495–504 (2014).

35. Meffre, E. et al. Surrogate light chain expressing human peripheral B cells produce self-reactive antibodies. *J. Exp. Med.* **199**, 145–150 (2004).

36. Amanat, F. et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* **26**, 1033–1036 (2020).

37. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2020).

38. Wang, Z. et al. Memory B cell responses to Omicron subvariants after SARS-CoV-2 mRNA breakthrough infection in humans. *J. Exp. Med.* **219**, e20221006 (2022).

39. Wang, Z. et al. Enhanced SARS-CoV-2 neutralization by dimeric IgA. *Sci. Transl. Med.* **13**, eabf1555 (2021).

40. Escolano, A. et al. Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* **570**, 468–473 (2019).

41. Viant, C., Escolano, A., Chen, S. T. & Nussenzweig, M. C. Sequencing, cloning, and antigen binding analysis of monoclonal antibodies isolated from single mouse B cells. *STAR Protoc.* **2**, 100389 (2021).

42. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

43. Soto, C. et al. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398–402 (2019).

44. Guo, Y., Chen, K., Kwong, P. D., Shapiro, L. & Sheng, Z. cAb-Rep: a database of curated antibody repertoires for exploring antibody diversity and predicting antibody prevalence. *Front. Immunol.* **10**, 2365 (2019).

Acknowledgements

We thank all of the study participants who devoted their time to our research; the Rockefeller University Hospital nursing staff and Clinical Research Support Office and nursing staff; all members of the M.C.N. laboratory for discussions; M. Jankovic and G. Scrivanti for laboratory support; T. Eisenreich for help with mouse colony management and technical help; and K. Gordon for cell sorting. This work was supported by NIH grants P01-AI138398-S1 and 2U19-AI111825 to M.C.N.; R37-AI64003 to P.D.B. and R01-AI78788 to T.H.; and 3UM1AI126620-5S1 to M. Caskey. D.S.-B. is supported in part by the National Center for Advancing Translational Sciences (NIH Clinical and Translational Science Award program, grant UL1-TR001866) and the Shapiro–Silverberg Fund for the Advancement of Translational Research. Z.W. was supported by the Maurice R. and Corinne P. Greenberg Center for the Advancement of Translational Research, in part by grant UL1 TR001866 from the National Center for Advancing Translational Sciences, NIH Clinical and Translational Science Award program. F.M. was supported by the Bulgari Women & Science Fellowship in COVID-19 Research. P.D.B. and M.C.N. are Howard Hughes Medical Institute (HHMI) Investigators. This Article is subject to HHMI's Open Access to Publications policy. HHMI laboratory heads have previously granted a non-exclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licences, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately on publication.

Author information

Author notes

1. These authors contributed equally: Dennis Schaefer-Babajew, Zijun Wang, Frauke Muecksch, Alice Cho

Authors and Affiliations

1. Laboratory of Molecular Immunology, The Rockefeller University, New York, NY, USA

   Dennis Schaefer-Babajew, Zijun Wang, Alice Cho, Maximilian Loewe, Melissa Cipolla, Raphael Raspe, Brianna Johnson, Victor Ramos, Martina Turroja, Katrina G. Millard, Juan Dizon, Irina Shimeliovich, Kai-Hui Yao, Thiago Y. Oliveira, Anna Gazumyan, Christian Gaebler, Marina Caskey & Michel C. Nussenzweig

2. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

   Frauke Muecksch, Marie Canis, Justin DaSilva, Fabian Schmidt, Leander Witte, Paul D. Bieniasz & Theodora Hatziioannou

3. Howard Hughes Medical Institute, New York, NY, USA

   Anna Gazumyan, Paul D. Bieniasz & Michel C. Nussenzweig

Contributions

D.S.-B. and M.C.N. conceptualized the study. D.S.-B., Z.W., F.M., A.C., P.D.B., T.H. and M.C.N. conceived, designed and analysed experiments. D.S.-B., C.G. and M. Caskey designed clinical protocols. D.S.-B., Z.W., F.M., A.C., M.L., M. Cipolla, R.R., M. Canis, J. DaSilva., F.S., L.W. and K.-H.Y. performed experiments. B.J. and A.G. produced antibodies. M.T., K.G.M., I.S., J. Dizon, C.G. and M. Caskey recruited participants, carried out clinical protocols and processed samples. V.R. and T.Y.O. performed bioinformatics analysis. D.S.-B. and M.C.N. wrote the manuscript with input from all of the authors.

Corresponding authors

Correspondence to Paul D. Bieniasz, Theodora Hatziioannou, Marina Caskey or Michel C. Nussenzweig.

Ethics declarations

Competing interests

The Rockefeller University has filed a provisional patent application in connection with this work, on which M.C.N. is listed as an inventor (US patent 63/021,387). The patent has been licensed by Rockefeller University to Bristol Meyers Squib. P.D.B. has received remuneration from Pfizer for consulting services related to SARS-CoV-2 vaccines.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 C135 and C144 – selectively abrogated binding to mutant RBDs and correlations with plasma reactivity.

a-**f**, Monoclonal antibody binding to mutant forms of RBD. **a**-**c**, Graphs show concentration-dependent antibody binding to (**a**) WT, (**b**) R436S/E484K, and (**c**) N440K/E484K RBDs by C144, C135, and Class 1 (C105), Class 2 (C952), Class 3 (C881), and Class 4 (C149)[5,6,8,26]. **d**-**f**,

Graphs show concentration dependent pre-pandemic healthy donor plasma binding to (**d**) WT, (**e**) R436S/E484K, and (**e**) N440K/E484K RBDs in the presence (purple) or absence (dotted lines) of 100 mg ml^{-1} of C135 and C144. Addition of C144 and C135 to plasma increases the binding activity of plasma against the WT but not the 2 mutant RBDs. **g**-**h**, Panels show the correlation of C135-LS (**g**) and C144-LS (**h**) serum levels at day 84 post-administration (around the times of vaccination as seen in Fig. 1b) with the total anti-WT RBD IgG antibody titre of mAb recipients (n = 18) after one (empty green circles) and two doses (solid green dots) of mRNA vaccination. Statistical significance in **g** ($r = 0.9097$ and $r = 0.5891$ with $p < 0.0001$ and $p = 0.0101$ for vax1 and vax2, respectively) and **h** ($r = 0.8772$ and $r = 0.5483$ with $p < 0.0001$ and $p = 0.0185$ for vax1 and vax2, respectively) was determined using the two-tailed Pearson correlation statistic. All experiments were performed at least in duplicate.

Extended Data Fig. 2 SARS-CoV-2 R346S/Q493K and R346S/N440K/E484K pseudotype virus neutralization by C135-LS and C144-LS, and BA.4/5 pseudotype neutralization by plasma.

a, Graphs show concentration-dependent neutralization curves for SARS-CoV-2 pseudoviruses by monoclonal antibodies. C144 (red), C135 (blue), and their equimolar combination (purple). **b**, Pre-pandemic plasma (squares) neutralization of WT or R346S/Q493K or R346S/N440K/E484K pseudoviruses in the absence or presence of 5 (purple dashed circles) or 100 µg ml^{-1} (purple solid circles) of C135 and C144[14]. **c**-**d**, As in (**b**) but for convalescent plasma with intermediate (**c**, COV157)[12] or strong (**d**, COV31)[12] neutralizing activity. The horizontal lines in all panels indicate half-maximal neutralization. **e**, Plasma half-maximal neutralizing titre (NT50s) against HIV-1 pseudotyped with the BA.4/5 S[38]. Each dot represents one individual from the control group (n = 31, blue) or from the mAb recipient group (n = 18, green). Red horizontal bars and red numbers represent median values. Statistical significance was determined using the two-tailed Mann-Whitney. For **a**-**d**, individual symbols represent the mean of two independent experiments and error bars the standard deviation. All experiments were performed at least in duplicate.

Extended Data Fig. 3 Flow-cytometry of human anti-RBD memory B cells.

a, Gating strategy for flow-cytometry phenotyping. Gating was on single lymphocytes that were $CD19^+$ and $CD20^+$, and $CD3^-$ $CD8^-$ $CD16^-$ Ova^- without uptake of live-dead dye (L/D). Antigen-specific cells were those with dual binding to Wuhan-Hu1 RBD-PE and RBD-AF647. Anti-IgG, -IgM were used to phenotype dual RBD-labelled B cells. **b**,**c**, Representative flow-cytometry plots of Wuhan Hu-1 RBD-binding memory B cells from mAb recipients after one and two doses of vaccination (**b**) and pre-pandemic health donors (**c**) serving as negative controls. Numbers in RBD-gate denote percentage of RBD dual-labelled cells of parent gate (see **a**). Corresponding flow-cytometry plots and gating strategy for vaccinated controls can be found in[11]. **d**-**e**, Number of IgG- (**d**) and IgM-expressing (**e**) WT RBD-specific memory B cells per 10 million $CD20^+$ B cells. Each dot represents one individual from the mAb recipient (green, n = 18) and control group (blue, n = 10)[9,11]. Horizontal red bars denote median values. **f**-**i**, Panels show the correlation of C135-LS (**f**, **g**) and C144-LS (**h**,**i**) serum levels at day 84 post-administration (around the times of vaccination as seen in Fig. 1b) with the percentage of WT RBD-specific memory B cells expressing either IgG (**f**, **h**) or IgM (**g**, **i**) after two vaccine doses as assessed by flow-cytometry. Green dots represent individual mAb recipients (n = 18). Statistical significance was determined using the two-tailed Mann-Whitney for **d** and **e**, and the Pearson r correlation statistic was used for **f**-**i**.

Extended Data Fig. 4 Fluorescence-activated cell sorting (FACS) of human anti-RBD memory B cells and subsequent BCR sequencing.

a-**b**, Panels showing IgG (**a**) and IgM (**b**) surface expression of anti-RBD memory cells exactly as in Fig. 2b and c, but with the 5 representative individuals from whom cells were subsequently sorted highlighted in yellow. **c**, Gating strategy for single-cell sorting of RBD-specific memory B cells. Dual-labelled (RBD-PE^+/-$AF647^+$) $CD20^+$ $CD3^-$ $CD8^-$ $CD16^-$ Ova^- cells were sorted. **d**, Representative flow cytometry plots show RBD-binding cells that were sorted from the 5 mAb recipients. **e**, Pie charts show the

distribution of antibody sequences derived from cells isolated from 10 vaccinated control individual after vax2. The upper panel shows IgM, and the lower panel depicts IgG sequences[9,11]. The number in the inner circle indicates the number of sequences analysed for the individual denoted above the circle. Slices coloured in shades of blue indicate cells that are clonally expanded (same IGHV and IGLV genes, with highly similar CDR3s). Pie slice size is proportional to the number of clonally related sequences. The black outline and % value indicate the frequency of clonally expanded sequences detected within an individual. White pie areas indicate the proportion of sequences isolated only once. For C005, there were no IgM transcripts amplified at the timepoint assayed. **f**-**h**, Comparison of the frequency distribution of V gene usage for the IgH and IgL among antibodies isolated from mRNA-vaccinated mAb recipients (this study) and controls[9,11], after vax2, and from database of shared clonotypes of human antibodies from Soto et al.[43]. Graphs show relative abundance of human IGHV (**f**), IGKV (**g**), and IGLV (**h**) genes within the human V gene database (in grey, Sequence Read Archive accession SRP010970), antibodies isolated from mAb recipients (in green) or vaccinated controls (in blue). Colours of stars indicate levels of statistical significance for the following frequency comparisons: black – vaccinated controls vs. database; red – mAb recipients vs. database; blue – mAb recipients vs. vaccinated controls.

Extended Data Fig. 5 WT RBD binding and WT SARS-Cov-2 neutralization by monoclonal pentameric IgM antibodies.

a-**b**, Panels depict WT RBD binding and WT SARS-CoV-2 S pseudotype neutralizing activity of a representative panels of monoclonal antibodies derived from IgM-expressing RBD-specific memory B cells expressed either as human IgG1 (IgG, as in Fig. 3a–c) or pentameric IgM (IgM[5]). **a**, Panel shows WT RBD EC50s of 15 monoclonal antibodies isolated from vaccinated mAb recipients (also see Supplementary Table 4). Grey shaded area between horizontal dotted lines indicates antibodies with EC50s >10 $\mu g\ ml^{-1}$ (poor binding) and non-binding antibodies arbitrarily grouped at 10 and 20 $\mu g\ ml^{-1}$, respectively. **b**, Plots show IC50s of 2 IgM-derived control antibodies (covering a wide range of neutralizing activity) in blue and 15 IgM-derived monoclonal antibodies from mAb recipients (as in **a**) in green, expressed as human IgG1 (IgG) or pentameric IgM (IgM[5]). For both panels

(**a**, **b**), ring plots summarize the fraction of antibodies in the indicated category among all tested (encircled number). Red horizontal bars and numbers indicate median values. For panel **a**, statistical significance was determined using the two-tailed Wilcoxon matched-pairs rank test to compare differences between the same monoclonal antibodies expressed as IgG or pentameric IgM, and the Chi-squared contingency statistic was used to compare categorical distributions from ring plots.

Extended Data Fig. 6 Competition BLI.

a-**d**, BLI traces of antibodies assayed for competition with class-reference antibodies. Traces show initial association curve (antigen capture phase of the primary antibody) and subsequent addition of secondary antibodies of unknown class. Thin solid black lines represent antibodies isolated from mAb recipients or vaccinated controls. Thick dashed lines are self-competition traces of C105 (green in **a**), C144 (red in **b**), C135 (blue in **c**) and C2172 (purple in **d**) for classes 1–4, respectively. **e**, Heat-map of relative inhibition of secondary antibody binding to the preformed capture antibody-RBD complexes (grey=no binding, red=unimpaired binding, orange=indeterminate). The left panel shows antibodies from mAb recipients, while the right panel shows IgM antibodies from vaccinated controls isolated in this study (both after vax2). Details on IgG antibodies isolated from vaccinated controls can be found in[9,11]. **f**, BLI traces defining C2172 as Class 4. C2172 is the primary/capture antibody (in dashed purple). The addition of known class-defining antibodies C105 (in green, Class 1[8]), C144 (in red, Class 2[8]), C135 (in blue, Class 3[8]), and C118 (in orange, Class 1/4[27]) establish C2172 as a bona fide Class 4 antibody.

Extended Data Fig. 7 Flow-cytometry of germinal center responses in mAb pre-treated mice and molecular characterization of germinal-centre-derived monoclonal antibodies.

a, Gating strategy for flow-cytometry phenotyping of germinal centre B cells isolated from the draining (popliteal) lymph nodes of WT C57BL/6 mice immunized with recombinant SARS-Cov-2 RBD 11 days prior (as

illustrated in Fig. 4a). Gating was on single lymphocytes that were B220$^+$ and CD4$^−$ CD8a$^−$ NK1.1$^−$ F4/80$^−$ (lineage-negative) without uptake of live-dead dye (L/D). Germinal centre B cells were those that were CD38$^−$ GL7$^+$ CD95 (Fas)$^+$. Antigen-specific cells were those with dual binding to Wuhan-Hu1 RBD-PE and RBD-AF647. **b**, Representative flow-cytometry plots of Wuhan Hu-1 RBD-binding germinal centre B cells from mice that had either received the combination of C135 and C144 (anti-RBD mAb group) or the irrelevant anti-HIV control antibodies 3BNC117 and 10–1074 (anti-HIV mAb controls) one day prior to immunization, with the percentage of binding cells denoted within the respective gate. **c**-**d**, Panels show the total number of CD38$^−$ GL7$^+$ CD95 (Fas)$^+$ germinal centre B cells (**c**) and RBD-binding germinal centre B cells (**d**) isolated from anti-RBD mAb pre-treated (n = 6, green) and control mice (n = 6, blue), with each dot corresponding to an individual mouse. **e**, Somatic hypermutation (SHM) levels of germinal centre B-cell-derived monoclonal antibodies shown as combined heavy- and light-chain variable region nucleotide substitutions plus one (IGVH+IGVL+1), with each dot representing one sequence from anti-RBD mAb pre-treated mice (green) or controls (blue). Ring plots below each column show the fraction of sequences with no (IGVH+IGVL+1 = 1) vs. any (IGVH+IGVL+1 > 1) SHM among all sequences analysed (encircled number) for the respective group. **f**, Percentage of sequences belonging to clones, defined as 2 or more sequences with the same IGHV and IGLV genes and with highly similar CDR3s, among all sequences obtained from the respective animal (as in Fig. 4d). Each dot represents one individual mouse from the anti-RBD mAb (n = 6, green) or control group (n = 6, blue). **g**, Affinity constants (K_d) of germinal centre B-cell-derived Fabs for WT SARS-CoV-2 RBD, as established from the monovalent interaction of Fabs with RBD monomers by BLI (also see Fig. 4f–i, Supplementary Table 6 and methods). Each dot represents a single Fab from the anti-RBD mAb (n = 8, green) or control group (n = 22, blue). Red horizontal bars (**c**-**g**) and numbers (**e**, **g**) indicate median (**c**, **d**, **f**, **g**) and mean (**e**) values. Statistical significance was determined using the two-tailed Mann-Whitney test for **c**-**g** **d**, and the two-sided Fisher's exact test was used to test the relative contribution of mutated and unmutated sequences in **e**.

Supplementary information

Supplementary Information

Descriptions for Supplementary Tables 1–6.

Reporting Summary

Supplementary Table 1

Cohort characteristics.

Supplementary Table 2

Individual participant characteristics.

Supplementary Table 3

Sequences of human anti-SARS-CoV-2 RBD antibodies.

Supplementary Table 4

Sequences, RBD ELISA binding, neutralization, affinity and epitopes of cloned recombinant human anti-SARS-CoV-2 RBD antibodies.

Supplementary Table 5

Sequences of mouse germinal centre B cell antibodies.

Supplementary Table 6

Sequences and binding affinities of cloned recombinant monoclonal mouse germinal centre Fab fragments.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation,

distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Schaefer-Babajew, D., Wang, Z., Muecksch, F. *et al.* Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination. *Nature* **613**, 735–742 (2023). https://doi.org/10.1038/s41586-022-05609-w

- Received: 04 June 2022
- Accepted: 30 November 2022
- Published: 06 December 2022
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05609-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **[Humoral immunity for durable control of SARS-CoV-2 and its variants]**
    - Ryutaro Kotaki
    - Saya Moriyama
    - Yoshimasa Takahashi

    Inflammation and Regeneration (2023)

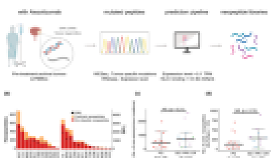
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05609-w



Blockade of the checkpoint receptor TIGIT prevents NK cell exhaustion and elicits potent anti-tumor immunity

18 June 2018

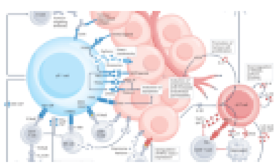
Qing Zhang, Jiacheng Bi, … Zhigang Tian



Neoantigen-specific CD8 T cell responses in the peripheral blood following PD-L1 blockade might predict therapy outcome in metastatic urothelial carcinoma

11 April 2022

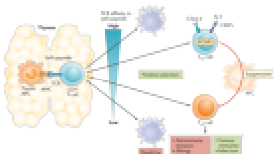
Jeppe Sejerø Holm, Samuel A. Funt, … Sine Reker Hadrup



The emerging roles of γδ T cells in cancer immunotherapy

09 January 2023

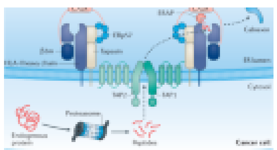
Sofia Mensurado, Rafael Blanco-Domínguez & Bruno Silva-Santos



Regulatory T cells in cancer immunosuppression — implications for anticancer therapy

31 January 2019

Yosuke Togashi, Kohei Shitara & Hiroyoshi Nishikawa



Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion

09 March 2021

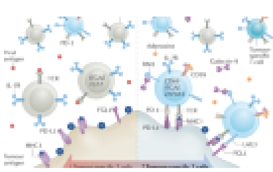
Suchit Jhunjhunwala, Christian Hammer & Lélia Delamarre



Translating gammadelta (γδ) T cells and their receptors into cancer cell therapies

06 September 2019

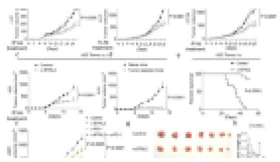
Zsolt Sebestyen, Immo Prinz, … Jurgen Kuball



Tumour antigen-induced T cell exhaustion — the archenemy of immune-hot malignancies

23 September 2021

Miguel Lopez de Rodas & Kurt A. Schalper



Cancer cell-expressed BTNL2 facilitates tumour immune escape via engagement with IL-17A-producing γδ T cells

11 January 2022

Yanyun Du, Qianwen Peng, … Chenhui Wang



Cytotoxic T cells isolated from healthy donors and cancer patients kill TGFβ-expressing cancer cells in a TGFβ-dependent manner

06 January 2021

Morten Orebo Holmström, Rasmus Erik Johansson Mortensen, … Mads Hald Andersen

- Article
- Open Access
- Published: 11 January 2023

γδ T cells are effectors of immunotherapy in cancers with HLA class I defects

- Natasja L. de Vries,
- Joris van de Haar,
- Vivien Veninga,
- Myriam Chalabi,
- Marieke E. Ijsselsteijn,
- Manon van der Ploeg,
- Jitske van den Bulk,
- Dina Ruano,
- Jose G. van den Berg,
- John B. Haanen,
- Laurien J. Zeverijn,
- Birgit S. Geurts,
- Gijs F. de Wit,
- Thomas W. Battaglia,
- Hans Gelderblom,
- Henk M. W. Verheul,
- Ton N. Schumacher,
- Lodewyk F. A. Wessels,
- Frits Koning,
- Noel F. C. C. de Miranda &
- …
- Emile E. Voest

Show authors

Nature **volume 613**, pages 743–750 (2023)

- 28k Accesses

- 267 Altmetric

- Metrics details

Abstract

DNA mismatch repair-deficient (MMR-d) cancers present an abundance of neoantigens that is thought to explain their exceptional responsiveness to immune checkpoint blockade (ICB)[1,2]. Here, in contrast to other cancer types[3,4,5], we observed that 20 out of 21 (95%) MMR-d cancers with genomic inactivation of β2-microglobulin (encoded by *B2M*) retained responsiveness to ICB, suggesting the involvement of immune effector cells other than $CD8^+$ T cells in this context. We next identified a strong association between *B2M* inactivation and increased infiltration by γδ T cells in MMR-d cancers. These γδ T cells mainly comprised the Vδ1 and Vδ3 subsets, and expressed high levels of PD-1, other activation markers, including cytotoxic molecules, and a broad repertoire of killer-cell immunoglobulin-like receptors. In vitro, $PD\text{-}1^+$ γδ T cells that were isolated from MMR-d colon cancers exhibited enhanced reactivity to human leukocyte antigen (HLA)-class-I-negative MMR-d colon cancer cell lines and *B2M*-knockout patient-derived tumour organoids compared with antigen-presentation-proficient cells. By comparing paired tumour samples from patients with MMR-d colon cancer that were obtained before and after dual PD-1 and CTLA-4 blockade, we found that immune checkpoint blockade substantially increased the frequency of γδ T cells in B2M-deficient cancers. Taken together, these data indicate that γδ T cells contribute to the response to immune checkpoint blockade in patients with HLA-class-I-negative MMR-d colon cancers, and underline the potential of γδ T cells in cancer immunotherapy.

Main

ICB targeting the PD-1–PD-L1 and/or CTLA-4 axes provides durable clinical benefits to patients who have cancers with MMR-d and high microsatellite instability[6,7,8,9]. The exceptional responses of cancers with MMR-d and high microsatellite instability to ICB is thought to be explained by their substantial burden of putative neoantigens, which originate from the extensive accumulation of mutations in their genomes[1,2]. This is consistent

with the current view that PD-1 blockade mainly boosts endogenous antitumour immunity driven by $CD8^+$ T cells, which recognize HLA-class-I-bound neoepitopes on cancer cells[10,11,12]. However, MMR-d colon cancers frequently lose HLA-class-I-mediated antigen presentation due to silencing of HLA class I genes, inactivating mutations in β2-microglobulin (encoded by *B2M*) or other defects in the antigen processing machinery[13,14,15,16], which can render these tumours resistant to $CD8^+$ T-cell-mediated immunity[3,4,5,17]. Notably, early evidence has indicated that B2M-deficient, MMR-d cancers can obtain durable responses to PD-1 blockade[18], suggesting that immune cell subsets other than $CD8^+$ T cells contribute to these responses.

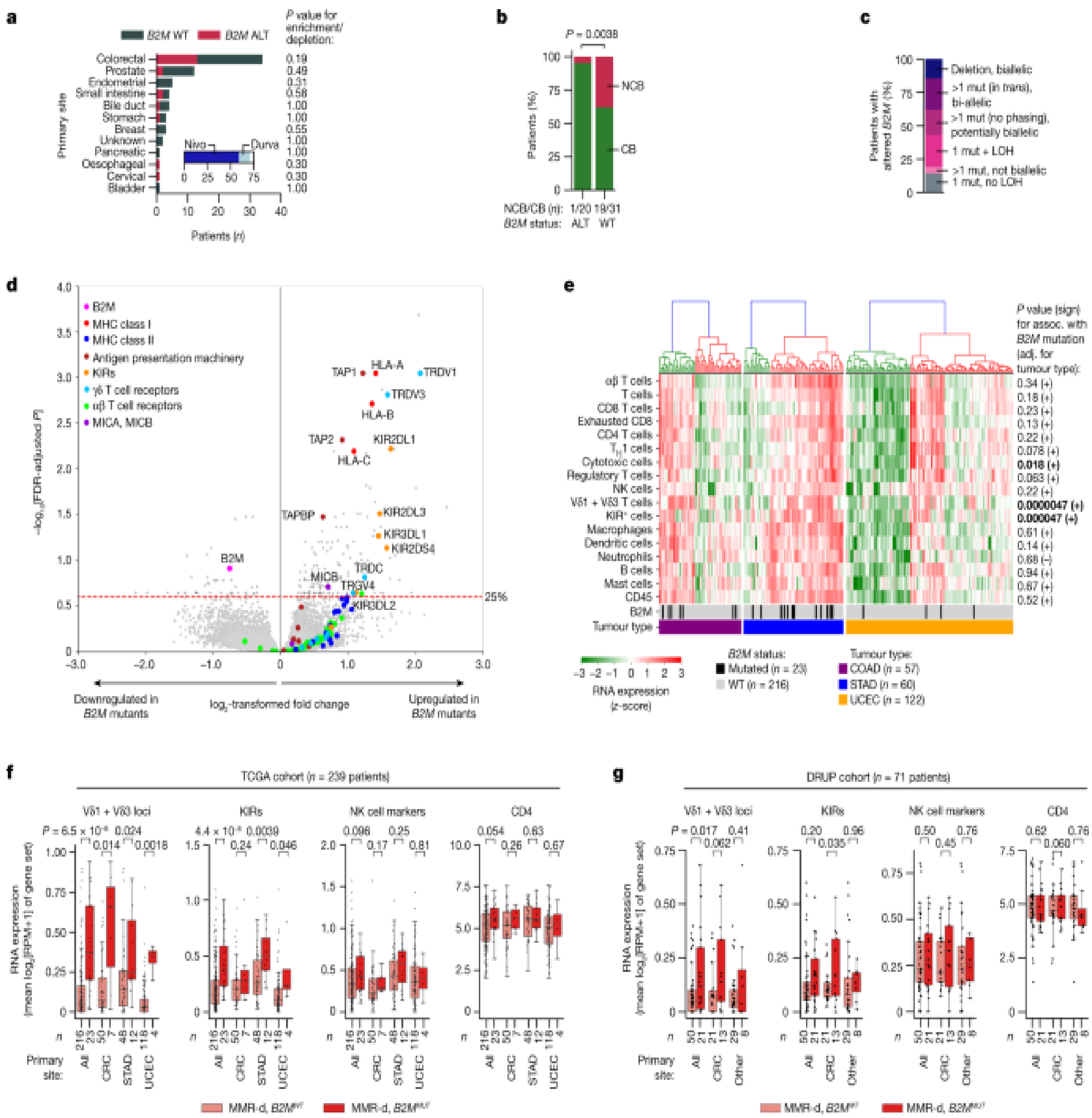
HLA-class-I-unrestricted immune cell subsets, which have the ability to kill tumour cells, include natural killer (NK) cells and γδ T cells. γδ T cells share many characteristics with their αβ T cell counterpart, such as cytotoxic effector functions, but express a distinct TCR that is composed of a γ and a δ chain. Different subsets of γδ T cells are defined by their TCR δ chain use, of which those expressing Vδ1 and Vδ3 are primarily 'tissue-resident' at mucosal sites, whereas those expressing Vδ2 are mainly found in blood[19]. Both adaptive and innate mechanisms of activation—for example, through stimulation of their γδ TCR or innate receptors such as NKG2D, DNAM-1, NKp30 or NKp44—have been described for γδ T cells[20]. Killer-cell immunoglobulin-like receptors (KIRs) are expressed by γδ T cells and regulate their activity depending on HLA class I expression in target cells[21]. Furthermore, γδ T cells were found to express high levels of PD-1 in MMR-d colorectal cancers (CRCs)[22], suggesting that these cells may be targeted by PD-1 blockade.

Here, we applied a combination of transcriptomic and imaging approaches for an in-depth analysis of ICB-naive and ICB-treated MMR-d colon cancers, as well as in vitro functional assays, and found evidence indicating that γδ T cells mediate responses to HLA-class-I-negative MMR-d tumours during treatment with ICB.

ICB is effective in *B2M* ^{MUT} MMR-d cancers

We evaluated responses to PD-1 blockade therapy in a cohort of 71 patients with MMR-d cancers from various anatomical sites treated in the Drug Rediscovery Protocol (DRUP)[23] in relation to their *B2M* status (Fig. 1a, Extended Data Fig. 1a–c and Supplementary Table 1). A clinical benefit (CB; defined as at least 4 months of disease control; the primary outcome of the DRUP) was observed in 20 out of 21 (95%) of patients with tumours with mutant or deleted *B2M* ($B2M^{MUT}$) tumours versus 31 out of 50 (62%) of patients with tumours with wild-type *B2M* ($B2M^{WT}$) (two-sided Fisher's exact test, $P = 0.0038$; logistic regression, $P = 0.022$ and $P = 0.027$, adjusted for tumour mutational burden (TMB), and TMB plus tumour type, respectively; Fig. 1b). Among patients with $B2M^{MUT}$ tumours, 3 out of 21 (14%) individuals experienced a complete response (according to RECIST1.1 criteria), 12 (57%) experienced a partial response, 5 (24%) experienced a durable stable disease and 1 (4.8%) experienced progressive disease as the best overall response. All 44 *B2M* alterations across 21 patients were clonal (Methods), consistent with previous observations in MMR-d cancers[18]. A total of 13 out of 21 (62%) patients with $B2M^{MUT}$ tumours had biallelic *B2M* alterations, 4 (19%) had potentially biallelic alterations and 4 (19%) had non-biallelic alterations (Fig. 1c and Methods). Non-biallelic alterations have also been associated with complete loss of B2M protein expression in MMR-d tumours[18]. Thus, *B2M* alterations are associated with a high clinical benefit rate of PD-1 blockade in patients with MMR-d cancers.

Fig. 1: In MMR-d cancers, *B2M* defects are positively associated with ICB responsiveness and infiltration by Vδ1 and Vδ3 T cells and KIR-expressing cells.



a, Tumour type distribution in the DRUP cohort (*n* = 71 patients). The colours denote patients' *B2M* status; grey, WT; red, altered (ALT). *P* values for the enrichment/depletion of *B2M*-altered tumours per primary site were calculated using two-sided Fisher's exact tests. The inset denotes the ICB treatment; dark blue, nivolumab (Nivo); light blue, durvalumab (Durva). **b**, *B2M* status (*x* axis) versus clinical benefit (green, CB; red, no clinical benefit (NCB)) of ICB treatment in the DRUP cohort. The *P* value was calculated using a two-sided Fisher's exact test. **c**, The allelic status of *B2M* alterations in the DRUP cohort. Mut, mutation. **d**, Differential gene expression between $B2M^{MUT}$ and $B2M^{WT}$ MMR-d cancers in the TCGA

COAD (colon adenocarcinoma; $n = 57$ patients), STAD (stomach adenocarcinoma; $n = 60$ patients) and UCEC (uterus corpus endometrial carcinoma; $n = 122$ patients) cohorts. The results were adjusted (adj.) for tumour type and multiple-hypothesis testing (Methods). **e**, Immune marker gene set expression in MMR-d cancers of the COAD, STAD and UCEC cohorts of the TCGA. The bottom two bars indicate *B2M* status and cancer type. The association (assoc.) between gene set expression and *B2M* status was tested using ordinary least squares linear regression (adjusted for tumour type; Methods), of which two-sided *P* values and the association sign are shown on the right. Cancers were ranked on the basis of hierarchical clustering (top dendrograms). *P* values less than 0.05 are in bold. **f**, Immune marker gene set expression in $B2M^{WT}$ (pink) and $B2M^{MUT}$ (red) MMR-d cancers in the TCGA COAD, STAD and UCEC cohorts separately or combined (all). Boxes, whiskers and dots indicate the quartiles, 1.5× the interquartile range (IQR) and individual data points, respectively. *P* values were calculated using two-sided Wilcoxon rank-sum tests. **g**, Immune marker gene set expression in $B2M^{WT}$ (pink) and $B2M^{MUT}$ (red) as described in **f**, but for MMR-d cancers in the DRUP cohort. Results are shown for all cancers combined, only CRC or all non-CRC cancers (other). Two-sided *P* values were calculated using linear regression, adjusting for biopsy site and tumour type (Methods).

Vδ1 and Vδ3 TCRs are overexpressed in $B2M^{MUT}$ cancers

To gain insights into the immune cell subsets that are involved in immune responses to HLA-class-I-negative MMR-d cancers, we used data of The Cancer Genome Atlas (TCGA) and studied the transcriptomic changes associated with the genomic loss of *B2M* in three cohorts of individuals with MMR-d cancer in colon adenocarcinoma (COAD; $n = 50$ ($B2M^{WT}$), $n = 7$ ($B2M^{MUT}$)), stomach adenocarcinoma (STAD; $n = 48$ ($B2M^{WT}$) and $n = 12$ ($B2M^{MUT}$)), and endometrium carcinoma (UCEC; $n = 118$ ($B2M^{WT}$) and $n = 4$ ($B2M^{MUT}$)). We found that *B2M* was among the most significantly downregulated genes in $B2M^{MUT}$ cancers (two-sided limma-voom-based regression, $P = 3.5 \times 10^{-4}$, Benjamini–Hochberg false-discovery rate (FDR)-

adjusted $P = 0.12$, adjusted for tumour type; Fig. 1d). Genes encoding components of the HLA class I antigen presentation machinery other than *B2M* were highly upregulated in $B2M^{MUT}$ tumours, which may reflect reduced evolutionary pressure on somatic inactivation of these genes in the $B2M^{MUT}$ context[18] (Fig. 1d). Notably, we found *TRDV1* and *TRDV3*, which encode the variable regions of the δ1 and δ3 chains of the γδ T cell receptor (TCR), among the most significantly upregulated loci in $B2M^{MUT}$ tumours (*TRDV1*, two-sided limma-voom-based regression, FDR-adjusted $P = 0.00090$, adjusted for tumour type; *TRDV3*, two-sided limma-voom-based regression, FDR-adjusted $P = 0.0015$, adjusted for tumour type; Fig. 1d), regardless of the allelic status of the *B2M* alteration (Extended Data Fig. 1d). Consistent with this, the expression levels of *TRDV1* and *TRDV3* were higher in $B2M^{MUT}$ compared with in $B2M^{WT}$ MMR-d cancers (two-sided Wilcoxon rank-sum test, $P = 6.5 \times 10^{-8}$ for all of the cohorts combined; two-sided linear regression, $P = 4.7 \times 10^{-6}$, adjusted for tumour type; Fig. 1d–f). Moreover, $B2M^{MUT}$ tumours showed overexpression of multiple KIRs (Fig. 1d), which clustered together with *TRDV1* and *TRDV3* on the basis of hierarchical clustering (Extended Data Fig. 1e). The expression level of different KIRs (Supplementary Table 2) was higher in $B2M^{MUT}$ tumours compared with in $B2M^{WT}$ MMR-d tumours (two-sided Wilcoxon rank-sum test, $P = 4.4 \times 10^{-6}$ for all cohorts combined; two-sided linear regression, $P = 4.7 \times 10^{-5}$, adjusted for tumour type; Fig. 1d–f). Together, these results suggest that ICB-naive $B2M^{MUT}$ MMR-d cancers show increased levels of Vδ1 and Vδ3 T cells as well as increased numbers of these or other immune cells expressing KIRs—a potential mechanism of recognition of HLA class I loss.

We used marker gene sets (modified from ref. [24]; Methods and Supplementary Table 2) to estimate the abundance of a broad set of other immune cell types on the basis of the RNA expression data of the TCGA cohorts. Hierarchical clustering identified a high- and a low-infiltrated cluster in each of the three tumour types (Fig. 1e). Compared with the Vδ1 and Vδ3 T cell and KIR gene sets, the other marker gene sets showed no or only weak association between expression level and *B2M* status, indicating that our findings were not solely driven by a generally more inflamed state of $B2M^{MUT}$ tumours (Fig. 1e,f and Extended Data Fig. 1f).

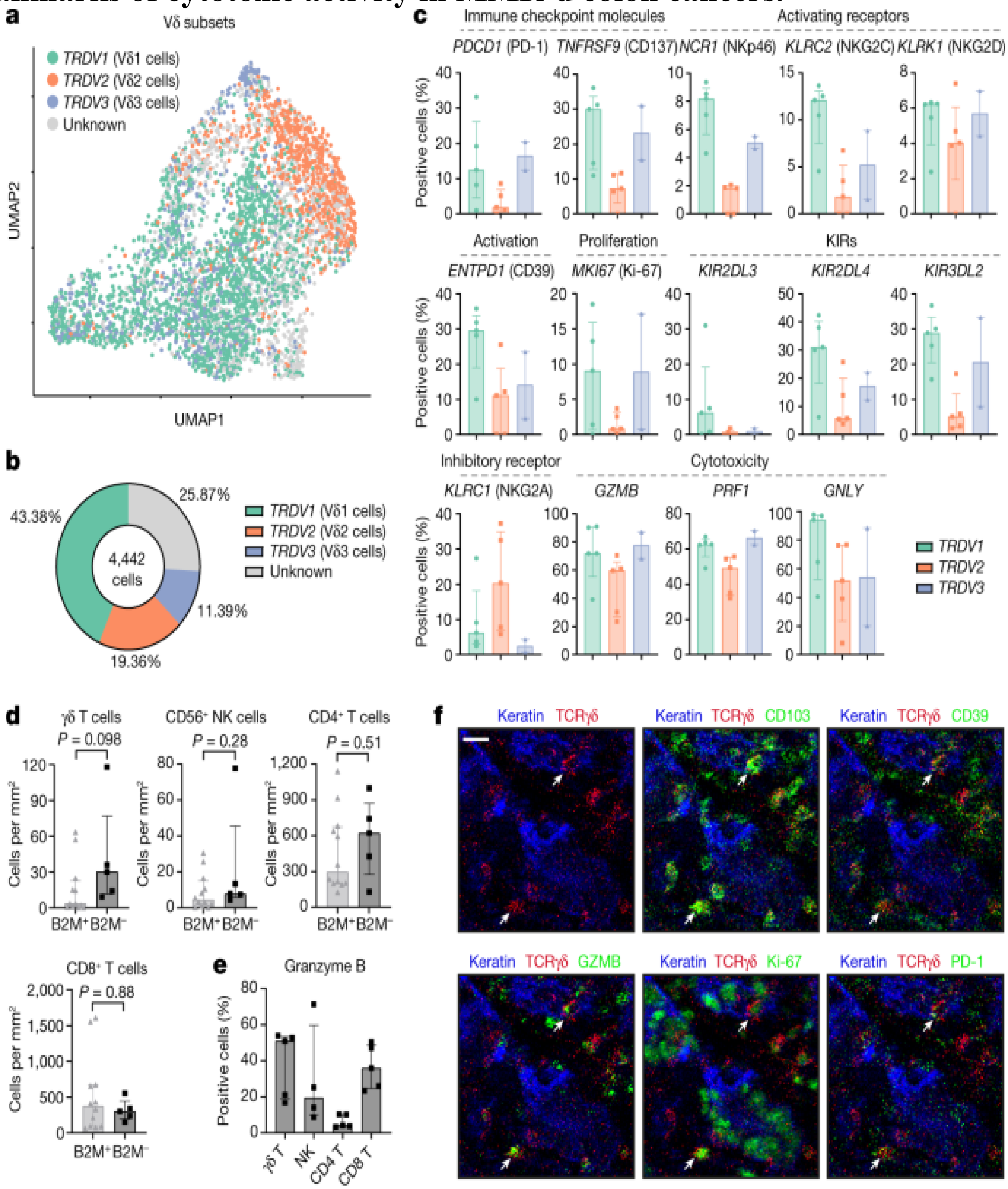
We next revisited the DRUP cohort and specifically applied the marker gene sets to RNA expression data. Despite the low patient numbers and high heterogeneity regarding tumour types and biopsy locations of this cohort, we confirmed increased *TRDV1* and *TRDV3* expression in $B2M^{MUT}$ tumours pan-cancer (two-sided linear regression, $P = 0.017$, adjusted for tumour type and biopsy site; Fig. 1g, Extended Data Fig. 1g and Methods). KIR expression was significantly associated with *B2M* status only in CRC (Fig. 1g). Across mismatch repair-proficient (MMR-p) metastatic cancers in the Hartwig database[25], 36 out of 2,256 (1.6%) cancers had a clonal *B2M* alteration, which was frequently accompanied by loss of heterozygosity (LOH) (Extended Data Fig. 1h and Supplementary Table 3). Although rare, *B2M* alterations were also significantly associated with increased expression of *TRDV1*/*TRDV3* loci in this context (two-sided linear regression, $P = 2.2 \times 10^{-17}$, adjusted for tumour type; Extended Data Fig. 1i and Methods). Taken together, *B2M* defects are positively associated with clinical benefits of ICB treatment, as well as infiltration by Vδ1 and Vδ3 T cells and expression of KIRs.

Vδ1 and Vδ3 T cells are activated in MMR-d CRC

To investigate which γδ T cell subsets are present in MMR-d colon cancers and to determine their functional characteristics, we performed single-cell RNA-sequencing (scRNA-seq) analysis of γδ T cells isolated from five MMR-d colon cancers (Extended Data Figs. 2 and 3 and Supplementary Table 4). Three distinct Vδ subsets were identified (Fig. 2a)—Vδ1 T cells were the most prevalent (43% of γδ T cells), followed by Vδ2 (19%) and Vδ3 T cells (11%) (Fig. 2b). *PDCD1* (encoding PD-1) was predominantly expressed by Vδ1 and Vδ3 T cells, whereas Vδ1 cells expressed high levels of genes that encode activation markers such as CD39 (*ENTPD1*) and CD38 (Fig. 2c and Extended Data Fig. 2b). Furthermore, proliferating γδ T cells (expressing *MKI67*) were especially observed in the Vδ1 and Vδ3 subsets (Fig. 2c). Other distinguishing features of the Vδ1 and Vδ3 T cell subsets included the expression of genes encoding activating receptors NKp46 (encoded by *NCR1*), NKG2C (encoded by *KLRC2*) and NKG2D (encoded by *KLRK1*) (Fig. 2c). Notably, the expression of several KIRs was also higher in the Vδ1 and Vδ3 subsets as compared to Vδ2 T cells (Fig. 2c). Almost all γδ T cells displayed expression of the genes encoding granzyme

B (*GZMB*), perforin (*PRF1*) and granulysin (*GNLY*) (Fig. 2c). Together, these data support a role for γδ T cells in mediating natural cytotoxic antitumour responses in HLA-class-I-negative MMR-d colon cancers.

Fig. 2: Tumour-infiltrating Vδ1 and Vδ3 T cell subsets display hallmarks of cytotoxic activity in MMR-d colon cancers.



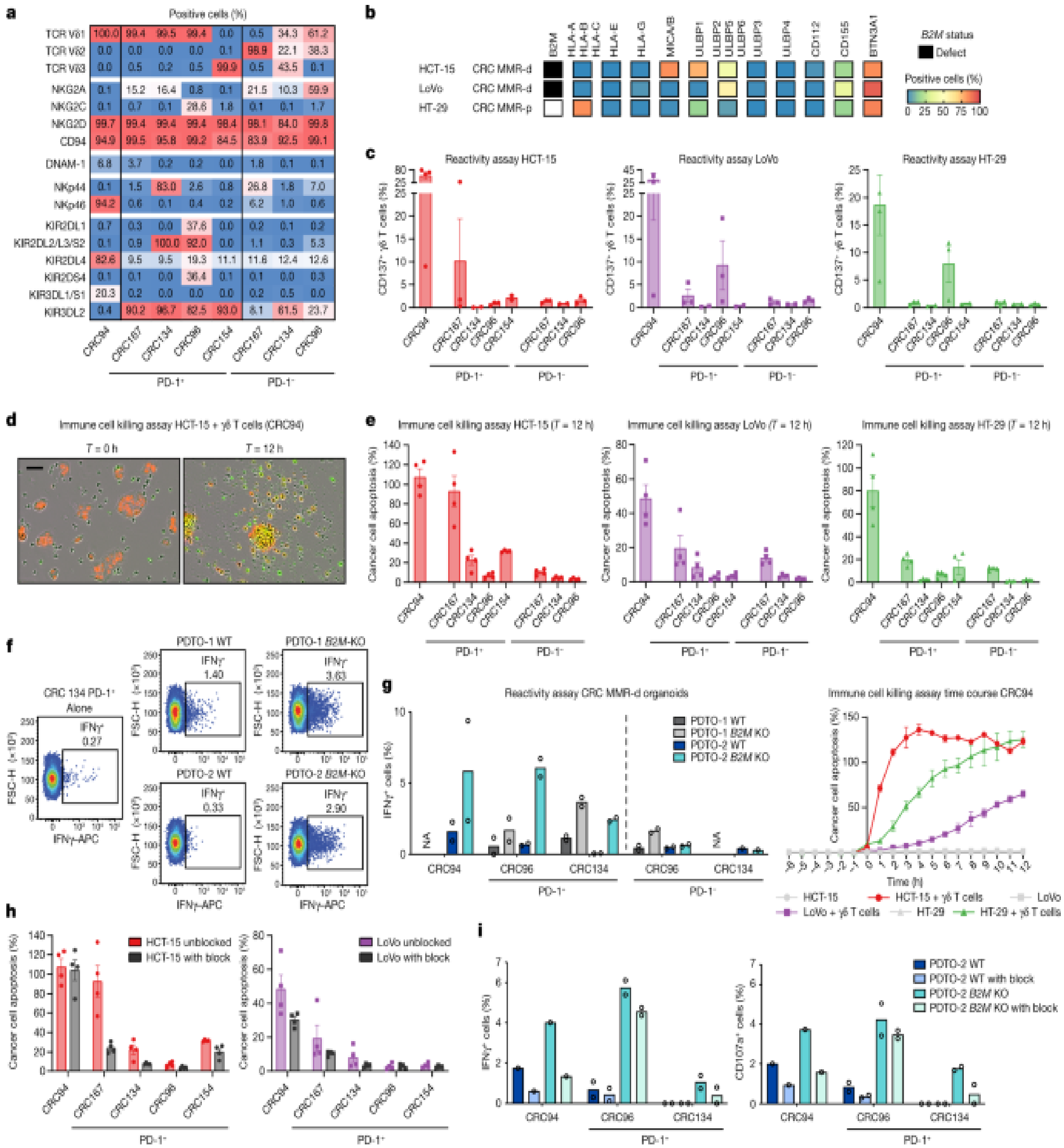
a, UMAP embedding showing the clustering of γδ T cells (*n* = 4,442) isolated from MMR-d colon cancers (*n* = 5) analysed using scRNA-seq. The colours represent the TCR Vδ chain usage. The functionally distinct γδ T cell clusters are shown in Extended Data Fig. 3. Dots represent single cells. **b**, The frequencies of TCR Vδ chain use of the γδ T cells (*n* = 4,442) analysed using scRNA-seq as a percentage of total γδ T cells. **c**, The frequencies of positive cells for selected genes across Vδ1 (*n* = 1,927), Vδ2 (*n* = 860) and Vδ3 (*n* = 506) cells as the percentage of total γδ T cells from each MMR-d colon tumour (*n* = 5) analysed using scRNA-seq. Vδ3 cells were present in two out of five colon cancers. Data are median ± IQR, with individual samples (dots). **d**, The frequencies of γδ T cells, $CD56^+$ NK cells, $CD4^+$ T cells and $CD8^+$ T cells in treatment-naive $B2M^+$ (*n* = 12) and $B2M^-$ (*n* = 5) MMR-d colon cancers. Data are median ± IQR, with individual samples (dots). *P* values were calculated using two-sided Wilcoxon rank-sum tests. **e**, The frequencies of granzyme-B-positive γδ T cells, $CD56^+$ NK cells, $CD4^+$ T cells and $CD8^+$ T cells in treatment-naive $B2M^-$ (*n* = 5) MMR-d colon cancers. $CD56^+$ NK cells were present in four out of five $B2M^-$ cancer samples. Data are median ± IQR, with individual samples (dots). **f**, Representative images of the detection of tissue-resident ($CD103^+$), activated ($CD39^+$), cytotoxic (granzyme B^+), proliferating (Ki-67^+) and PD-1^+ γδ T cells (white arrows) by IMC analysis of a treatment-naive MMR-d colon cancer with B2M defects. Scale bar, 20 μm.

Next, we applied imaging mass cytometry (IMC) analysis to a cohort of 17 individuals with ICB-naive MMR-d colon cancers (Supplementary Table 4). High levels of γδ T cell infiltration were observed in cancers with B2M defects as compared to B2M-proficient cancers, albeit this difference was not significant (Fig. 2d). The levels of other immune cells, including NK cells, $CD4^+$ T cells and $CD8^+$ T cells, were similar between B2M-deficient and B2M-proficient tumours (Fig. 2d). In B2M-deficient cancers, γδ T cells showed frequent intraepithelial localization and expression of CD103 (tissue-residency), CD39 (activation), granzyme B (cytotoxicity) and Ki-67 (proliferation), as well as PD-1 (Fig. 2d–f and Extended Data Fig. 2c), consistent with the scRNA-seq data. Notably, γδ T cells in B2M-deficient cancers showed co-expression of CD103 and CD39 (Extended Data Fig. 2d), which has been reported to identify tumour-reactive $CD8^+$ αβ T cells in a variety of cancers[26].

PD-1^{+} Vδ1 and Vδ3 T cells kill HLA-class-I^{-} CRC cells

We next sought to determine whether tumour-infiltrating γδ T cells can recognize and kill CRC cells. We isolated and expanded PD-1^{-} and PD-1^{+} γδ T cells from five MMR-d colon cancers (Extended Data Fig. 4a–c and Supplementary Table 4). Consistent with the scRNA-seq data, expanded PD-1^{+} γδ T cell populations lacked Vδ2^{+} cells and comprised the Vδ1^{+} or Vδ3^{+} subsets, whereas the PD-1^{-} fractions contained Vδ2^{+} or a mixture of Vδ1^{+}, Vδ2^{+} and Vδ3^{+} populations (Fig. 3a and Extended Data Fig. 4d). Detailed immunophenotyping of the expanded γδ T cells (Fig. 3a and Extended Data Fig. 5a) showed that all of the subsets expressed the activating receptor NKG2D, whereas the surface expression of natural cytotoxicity receptors (NCRs) and KIRs was most frequent on PD-1^{+} γδ T cells (Vδ1 or Vδ3^{+}), consistent with the scRNA-seq results of unexpanded populations.

Fig. 3: γδ T cells from MMR-d colon cancers show preferential reactivity to HLA-class-I-negative cancer cell lines and organoids.



a, The percentage of positive cells for the indicated markers on expanded γδ T cells from MMR-d colon cancers ($n = 5$). **b**, Diagram showing the *B2M* status and surface expression of HLA class I, NKG2D ligands, DNAM-1 ligands and butyrophilin on CRC cell lines. MMR-p, MMR proficient. **c**, CD137 expression on γδ T cells after co-culture with CRC cell lines. Data are mean ± s.e.m. from at least two independent experiments. **d**, Representative images showing the killing of NucLightRed-transduced HCT-15 cells by γδ T cells in the presence of a green fluorescent caspase-3/7 reagent. Cancer cell apoptosis is visualized in yellow. Scale bar, 50 μm. **e**,

Quantification of the killing of CRC cell lines after co-culture with γδ T cells as described in **d**. Data are mean ± s.e.m. of two wells with two images per well. A representative time course of cancer cell apoptosis is shown at the bottom right. **f**, Representative flow cytometry plots showing IFNγ expression in γδ T cells unstimulated (alone) and after stimulation with two *B2M*^{WT} and *B2M*^{KO} CRC MMR-d organoids. **g**, IFNγ expression in γδ T cells after stimulation with two *B2M*^{WT} and *B2M*^{KO} CRC MMR-d organoids, shown as the difference compared with the unstimulated γδ T cell sample. Data are from two biological replicates, except for a single biological replicate of CRC134 PD-1^{-}. NA, not available. **h**, The killing of CRC cell lines after 12 h co-culture with γδ T cells with or without NKG2D ligand blocking. Data are mean ± s.e.m. of two wells with two images per well. **i**, IFNγ (left) and CD107a (right) expression in γδ T cells after stimulation with *B2M*^{WT} PDTO-2 or *B2M*^{KO} PDTO-2, with or without NKG2D ligand blocking and subtracted background signal. Data are from two biological replicates, except for a single biological replicate of CRC94.

We measured the reactivity of the expanded γδ T cell populations to HLA-class-I-negative and HLA-class-I-positive cancer cell lines (Fig. 3b and Extended Data Fig. 4b). After co-culture with the different cancer cell lines, reactivity (assessed by expression of activation markers and secretion of IFNγ) was largely restricted to PD-1^{+} γδ T cells (Vδ1 or Vδ3^{+}), whereas activation of PD-1^{-} γδ T cells (Vδ2^{+}) was generally not detected (Fig. 3c and Extended Data Fig. 4). PD-1^{+} γδ T cell (Vδ1 or Vδ3^{+}) reactivity was variable and was observed against both HLA-class-I-negative and HLA-class-I-positive cell lines (Fig. 3c and Extended Data Fig. 4). To quantify and visualize the differences in the killing of CRC cell lines by PD-1^{+} and PD-1^{-} γδ T cells, we co-cultured the γδ T cell populations with three CRC cell lines (HCT-15, LoVo, HT-29) in the presence of a fluorescent cleaved-caspase-3/7 reporter to measure cancer cell apoptosis over time (Fig. 3d,e). We found pronounced cancer cell apoptosis after co-culture with PD-1^{+} γδ T cells (Vδ1 or Vδ3^{+}) compared with PD-1^{-} cells; cancer cell death was more pronounced in HLA-class-I-negative HCT-15 cells (Fig. 3e and Supplementary Videos 1 and 2). Reintroduction of *B2M* in the *B2M*-deficient HCT-15 and LoVo cells diminished their killing by PD-1^{+} γδ T

cells (Vδ1 or Vδ3$^+$) cells (Extended Data Fig. 6), suggesting that *B2M* loss increases the sensitivity to γδ T cells.

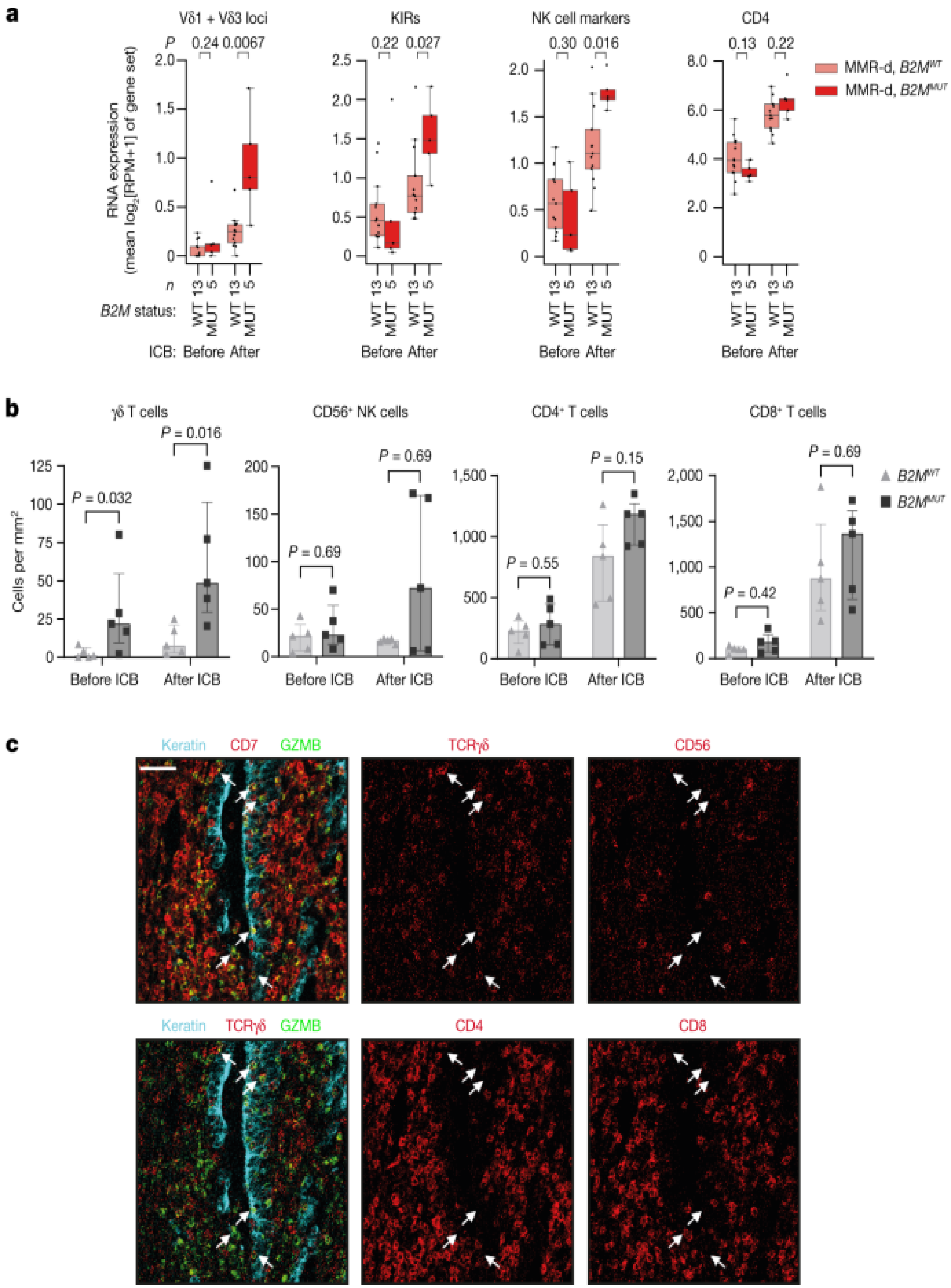
Next, we established two parental patient-derived tumour organoid lines (PDTOs; Supplementary Table 5) of MMR-d CRC and generated isogenic *B2M*^{KO} lines using CRISPR. Genomic knockout of *B2M* effectively abrogated cell surface expression of HLA class I (Extended Data Fig. 7). We exposed two *B2M*^{KO} lines and their parental *B2M*^{WT} lines to the expanded γδ T cell subsets, and quantified γδ T cell activation by determination of IFNγ expression. Similar to our cell line data, γδ T cells displayed increased reactivity to *B2M*^{KO} PDTOs in comparison to the *B2M*^{WT} PDTOs (Fig. 3f,g). Furthermore, γδ T cell reactivity to *B2M*^{KO} tumour organoids was preferentially contained within the PD-1$^+$ population of γδ T cells (Fig. 3g). Thus, a lack of HLA class I antigen presentation in MMR-d tumour cells can be effectively sensed by γδ T cells and stimulates their antitumour response.

Expression of NKG2D on γδ T cells decreased during co-culture with target cells (Extended Data Fig. 8a,b), suggesting the involvement of the NKG2D receptor in γδ T cell activity. The NKG2D ligands MICA/B and ULBPs were expressed by the cancer cell lines (Fig. 3b) and the MMR-d CRC PDTOs, irrespective of their *B2M* status (Extended Data Fig. 7). To examine which receptor–ligand interactions might regulate the activity of PD-1$^+$ γδ T cells, we performed blocking experiments focused on (1) NKG2D, (2) DNAM-1 and (3) γδ TCR signalling. Of these candidates, the only consistent inhibitory effect was observed for NKG2D ligand blocking on cancer cells, which decreased the activation and killing ability of most PD-1$^+$ γδ T cells (Fig. 3h and Extended Data Fig. 8c,d), confirming the mechanistic involvement of the NKG2D receptor in γδ T cell activation in this context. Moreover, blocking NKG2D ligands on MMR-d CRC PDTOs reduced the PDTO-directed tumour reactivity of γδ T cells from CRC94 and CRC134 (Fig. 3i). Together, these results show that γδ T cell reactivity to MMR-d tumours is partly dependent on NKG2D/NKG2D-ligand interactions.

ICB boosts Vδ1 and Vδ3 T cells in *B2M*^{MUT} CRC

We subsequently studied how ICB influences γδ T cell infiltration and activation in MMR-d colon cancers in the therapeutic context. For this purpose, we analysed pre- and post-treatment samples of the NICHE trial[9], in which patients with colon cancer were treated with neoadjuvant PD-1 plus CTLA-4 blockade. Consistent with our observations in the DRUP cohort, 4 out of 5 (80%) individuals with $B2M^{MUT}$ cancers in the NICHE trial showed a complete pathologic clinical response. Immunohistochemical analysis confirmed the loss of B2M protein expression on tumour cells in all mutated cases (Extended Data Fig. 9). Whereas expression of immune marker gene sets in the pretreatment samples was similar between 5 $B2M^{MUT}$ versus 13 $B2M^{WT}$ cancers, ICB induced a clear immunological divergence between these two groups (Fig. 4a). The $B2M^{MUT}$ subgroup was most significantly associated with higher post-treatment expression of *TRDV1* and *TRDV3* (two-sided Wilcoxon rank-sum test, $P = 0.0067$; Fig. 4a), followed by higher expression of the general immune cell marker CD45, NK-cell-related markers, KIRs and αβTCRs (two-sided Wilcoxon rank-sum test, $P = 0.016$, $P = 0.016$, $P = 0.027$ and $P = 0.043$, respectively; Fig. 4a and Extended Data Fig. 10a). The set of KIRs upregulated after ICB in $B2M^{MUT}$ cancers (Extended Data Fig. 10b) was consistent with the sets of KIRs upregulated in $B2M^{MUT}$ MMR-d cancers in TCGA (Fig. 1e), and those expressed by MMR-d tumour-infiltrating γδ T cells (Fig. 2c). Pre- and post-ICB gene expression levels related to CD4 and CD8 infiltration were not associated with *B2M* status (Fig. 4a and Extended Data Fig. 10a).

Fig. 4: ICB induces substantial infiltration of γδ T cells into MMR-d colon cancers with defects in antigen presentation.



a, The RNA expression of different immune marker gene sets in MMR-d *B2M*^{WT} (pink) and MMR-d *B2M*^{MUT} (red) cancers before (left) and after

(right) neoadjuvant ICB in the NICHE study. The boxes, whiskers and dots indicate quartiles, 1.5 × IQR and individual data points, respectively. *P* values were calculated using two-sided Wilcoxon rank-sum tests comparing MMR-d $B2M^{WT}$ versus MMR-d $B2M^{MUT}$ cancers. **b**, The frequencies of γδ T cells, $CD56^{+}$ NK cells, $CD4^{+}$ T cells and $CD8^{+}$ T cells in $B2M^{WT}$ ($n = 5$) and $B2M^{MUT}$ ($n = 5$) MMR-d colon cancers before and after ICB treatment. Data are median ± IQR, with individual samples (dots). *P* values were calculated using two-sided Wilcoxon rank-sum tests. **c**, Representative images of granzyme-B-positive γδ T cells infiltrating the tumour epithelium (white arrows) by IMC analysis of a $B2M^{MUT}$ MMR-d colon cancer after ICB treatment. Scale bar, 50 μm.

To quantify and investigate the differences in immune profiles after ICB treatment, we used IMC to analyse tissues derived from five $B2M^{MUT}$ HLA-class-I-negative and five $B2M^{WT}$ HLA-class-I-positive cancers before and after ICB treatment. In the ICB-naive setting, $B2M^{MUT}$ MMR-d colon cancers showed higher γδ T cell infiltration compared with $B2M^{WT}$ MMR-d colon cancers (two-sided Wilcoxon rank-sum test, $P = 0.032$; Fig. 4b and Extended Data Fig. 10c). Importantly, a large proportion of these γδ T cells showed an intraepithelial localization in $B2M^{MUT}$ MMR-d colon cancers compared with the $B2M^{WT}$ samples (two-sided Wilcoxon rank-sum test, $P = 0.0079$; Extended Data Fig. 10d). No significant differences were observed in the infiltration of other immune cells, such as NK cells, $CD4^{+}$ T cells and $CD8^{+}$ T cells, in ICB-naive $B2M^{MUT}$ versus $B2M^{WT}$ MMR-d colon cancers (Fig. 4b). ICB treatment resulted in major pathologic clinical responses, and residual cancer cells were absent in most post-ICB samples. All post-ICB tissues showed a profound infiltration of different types of immune cells (Extended Data Fig. 10e), of which γδ T cells were the only immune subset that was significantly higher in ICB-treated $B2M^{MUT}$ compared with $B2M^{WT}$ MMR-d colon cancers (two-sided Wilcoxon rank-sum test, $P = 0.016$; Fig. 4b and Extended Data Fig. 10c). In the sole $B2M^{MUT}$ case that still contained cancer cells after treatment with ICB, the majority of granzyme B^{+} immune cells infiltrating the tumour epithelium were γδ T cells (Fig. 4c). These γδ T cells displayed co-expression of CD103, CD39, Ki-67 and PD-1 (Extended Data Fig. 10f–h). Taken together, these results show that ICB treatment of MMR-d colon cancer increases the presence of activated, cytotoxic and

proliferating γδ T cells at the tumour site, especially when these cancers are B2M-deficient, highlighting γδ T cells as effectors of ICB treatment within this context.

Discussion

$CD8^+$ αβ T cells are major effectors of ICB[11,12,27] and rely on HLA class I antigen presentation of target cells. We confirm and shed light on the paradox that patients with HLA class I defects in MMR-d cancers retain the clinical benefit of ICB, suggesting that other immune effector cells are involved in compensating for the lack of conventional $CD8^+$ T cell immunity in this setting. We show that genomic inactivation of *B2M* in MMR-d colon cancers was associated with: (1) an elevated frequency of activated γδ T cells in ICB-naive tumours; (2) an increased presence of tumour-infiltrating γδ T cells after ICB treatment; (3) in vitro activation of tumour-infiltrating γδ T cells by CRC cell lines and PDTOs; and (4) killing of tumour cell lines by γδ T cells, in particular by Vδ1 and Vδ3 subsets expressing PD-1.

Different subsets of γδ T cells exhibit substantially diverse functions that, in the context of cancer, range from tumour-promoting to tumouricidal effects[20,28,29]. Thus, it is of interest to determine what defines antitumour reactivity of γδ T cells. Here we isolated Vδ1/3-expressing $PD\text{-}1^+$ T cells as well as Vδ2-expressing $PD\text{-}1^-$ T cells from MMR-d tumour tissues. Our data suggest that especially tumour-infiltrating Vδ1 and Vδ3 T cells can recognize and kill HLA-class-I-negative MMR-d tumours, whereas Vγ9Vδ2 cells, the most studied and main subset of γδ T cells in the blood, appear to be less relevant within this context. This is consistent with other studies showing that the cytotoxic ability of Vδ1 cells generally outperforms their Vδ2 counterparts[30,31,32,33,34]. Notably, reports of the cytotoxicity of tumour-infiltrating Vδ3 cells have been lacking. Furthermore, the observation that $PD\text{-}1^+$ γδ T cells (Vδ1 and Vδ3 phenotype) demonstrated clearly higher levels of antitumour reactivity compared with their $PD\text{-}1^-$ counterparts (Vδ2 phenotype) suggests that, as for $CD8^+$ αβ T cells[35], PD-1 expression may be a marker of antitumour reactivity in γδ T cells.

The mechanisms of activation of γδ T cells are notoriously complex and diverse[20]. Specifically, for Vδ1^{+} cells, NKG2D has been described to be involved in tumour recognition, which is dependent on tumour cell expression of NKG2D ligands MICA/B and ULBPs[36,37,38]. Here, MICA/B and ULBPs were highly expressed by the MMR-d CRC cell lines and tumour organoids, and blocking these ligands reduced γδ T cell activation and cytotoxicity. This suggests a role for the activating receptor NKG2D in γδ T cell immunity to MMR-d tumours. Future research should address the outstanding question of how γδ T cells accumulate in B2M-deficient tumours, and whether the lack of CD8^{+} T cell activity might contribute to the establishment of an attractive niche for γδ T cells and other immune effector cells. Potential mechanisms for the recognition of HLA-class-I-negative phenotypes may include KIR-, NKG2A- and LILRB1-mediated interactions with target cancer cells. Notably, we found that the expression of KIRs was most pronounced on PD-1^{+} γδ T cells (Vδ1 or Vδ3^{+} subsets), which demonstrated anti-tumour activity. Whether the lack of KIR-mediated signalling promotes the survival of γδ T cells and their intratumoural proliferation remains to be studied.

Our findings have broad implications for cancer immunotherapy. First, our findings strengthen the rationale for combining PD-1 blockade with immunotherapeutic approaches to further enhance γδ T-cell-based antitumour immunity. Second, the presence or absence in tumours of specific γδ T cell subsets (such as Vδ1 or Vδ3) may help to define patients who are responsive or unresponsive to ICB, respectively, especially in the case of MMR-d cancers and other malignancies with frequent HLA class I defects, such as stomach adenocarcinoma[39] and Hodgkin's lymphoma[40]. Third, our results suggest that MMR-d cancers and other tumours with HLA class I defects may be particularly attractive targets for Vδ1 or Vδ3 T-cell-based cellular therapies.

Although we have provided detailed and multidimensional analyses, it is probable that γδ T cells are not the only factor driving ICB responses in HLA-class-I-negative MMR-d CRC tumours. In this context, other HLA-class-I-independent immune subsets, such as NK cells and neoantigen-specific CD4^{+} T cells may also contribute. The latter were shown to have an important role in the response to ICB (as reported in mouse *B2M*-deficient

MMR-d cancer models[41]), and may also support γδ T-cell-driven responses. Notably, no subset equivalent to Vδ1 or Vδ3 T cells has been identified in mice, which complicates their investigation in in vivo models. In conclusion, our results provide strong evidence that γδ T cells are cytotoxic effector cells of ICB treatment in HLA-class-I-negative MMR-d colon cancers, with implications for further exploitation of γδ T cells in cancer immunotherapy.

Methods

TCGA data

RNA expression data (raw counts) of the colon adenocarcinoma (COAD), stomach adenocarcinoma (STAD) and Uterus Corpus Endometrium Carcinoma (UCEC) cohorts of The Cancer Genome Atlas (TCGA) Research Network were downloaded through the GDC data portal (https://portal.gdc.cancer.gov) on 10 April 2019. Of these cohorts, mutation, copy number, purity and ploidy data were downloaded from the GDC on 11 November 2021, as the controlled access ABSOLUTE-annotated[42] MAF file (mutations), SNP6 white-listed copy number segments file (copy numbers) and ABSOLUTE purity/ploidy file of the TCGA PanCanAtlas project[43]. Mismatch-repair-deficiency status was obtained from ref. [44] (TCGA subtype = GI.HM-indel or UCEC.MSI).

DRUP data

A detailed description of the DRUP, including details on patient accrual, study design, oversight and end points was published previously[23]. In brief, the DRUP is a national, non-randomized multidrug and multitumour study in the Netherlands, in which patients receive off-label drugs registered for other treatment indications. These patients had advanced or metastatic solid tumours and had exhausted standard-treatment options, they were required to be at least 18 years of age, with acceptable organ function and performance status (Eastern Cooperative Oncology Group (ECOG) score ≤ 2), and have an objectively evaluable disease of which a fresh baseline tumour biopsy could safely be obtained. We analysed 71 patients with MMR-d cancers recruited and treated with PD-1 blockade in 22 Dutch

hospitals participating in the DRUP[23] between 2016 and 2021. Patients included in this analysis had (1) a clinical follow-up ≥16 weeks after start of PD-1 blockade treatment; (2) WGS data passing standard quality controls (as defined previously, including a sequencing-based tumour purity of ≥20%)[25]; and (3) available RNA-seq data (Supplementary Table 1). MMR-d status was determined using routine diagnostics at the hospital of patient accrual and was confirmed by WGS, on the basis of an MSIseq (v.1.0.0)[45] score of >4, which represents a predefined threshold[25]. Consistent with the study protocol of the DRUP[23], the primary outcome measure for our analysis was clinical benefit, defined as disease control of ≥16 weeks, and the secondary outcome measure was best overall response, all assessed according to the RECIST 1.1 guidelines by the local treatment team at the site of accrual. As determined in the study protocol, these outcome measures were considered to be evaluable in patients who received at least two cycles of intravenous study medication, and for whom the response was radiologically or clinically evaluable (at the treating physician's discretion). For genomics and transcriptomics analyses, fresh frozen tumour biopsies were obtained at the baseline (that is, before PD-1 blockade). WGS analysis (median depths, ~100× and ~40× for tumour and normal, respectively) and bioinformatics analyses were performed as previously described[23,25], with an optimized pipeline based on open-source tools that is freely available at GitHub (https://github.com/hartwigmedical/pipeline5). The TMB per Mb was determined by counting the genome-wide number of mutations (SNVs, MNVs and indels) and dividing this number by the number of megabases sequenced. For RNA-seq analysis, we extracted total RNA using the QIAGEN QIAsymphony RNA kit (931636). Samples with approximately 100 ng total RNA were prepared using KAPA RNA Hyper + RiboErase HMR (8098131702) and the RNA libraries were paired-end sequenced on the Illumina NextSeq550 platform (2 × 75 bp) or the Illumina NovaSeq6000 platform (2 × 150 bp). Raw RNA reads (FASTA files) were aligned to the human reference genome (GRCh38) using STAR[46], (v.2.7.7a) using the default settings in two-pass mode.

Hartwig data

We analysed 2,256 metastatic tumours included in the freely available Hartwig database[25] that (1) were MMR-p (WGS-based MSIseq[45] score ≤4);

(2) had available WGS data passing standard quality controls (as defined before, including a sequencing-based tumour purity ≥20%)[25]; (3) and had available RNA-seq data. We excluded 89 tumours from rare primary tumour locations, defined as locations with less that <20 patients in our selection. When individual patients had data available of biopsies obtained at different timepoints, we included data of only the first biopsy. Sequencing and bioinformatics were performed identically to the procedures used for the generation of the DRUP dataset (see above). Details of this cohort are provided in Supplementary Table 3.

NICHE study data

Raw RNA reads (FASTA files) of our recently published NICHE study[9] (ClinicalTrials.gov: NCT03026140) were generated as described in the original publication and aligned to the human reference genome (GRCh38) with STAR[46] (v.2.7.7a) using the default settings in two-pass mode. For gene expression quantification, we used the gencode.v35.annotation.gtf annotation file. Somatic mutation data were obtained from DNA-seq of pretreatment tumour biopsies and matched germline DNA, as described in the original publication[9].

B2M status

Consistent with the notion that both biallelic and monoallelic non-synonymous *B2M* mutations are strongly associated with tumour-specific loss of B2M protein expression[18], we considered all tumours with at least one somatic, non-synonymous *B2M* mutation to be a *B2M* mutant. As none of the *B2M* mutant tumours in TCGA had *B2M* copy number gains or losses, LOH of *B2M* could easily be assessed by a simple calculation estimating the mutation's copy number:

$${{\rm{Mut}}}_{{\rm{CN}}}={\rm{round}}\left(2\times \frac{{\rm{VAF}}}{{\rm{purity}}}\right)$$

where $\mathrm{Mut}_{\mathrm{CN}}$ represents the estimated mutation's copy number (rounded to an integer value), VAF represents the variant allele frequency of the

mutation and purity equals the ABSOLUTE-based[42] tumour cell fraction of the sample.

A Mut_{CN} equal to 2 was considered to be consistent with LOH, as the most parsimonious explanation of such a result is the scenario in which all tumour-derived reads spanning the region of the *B2M* mutation contain the mutation and none of the tumour-derived reads are WT.

In analyses of patients in the DRUP and Hartwig datasets, LOH of *B2M* mutations was determined as an integrated functionality of PURPLE (v.2.34)[47]. When multiple *B2M* mutations were present within a sample, we manually phased the mutations through inspection of the *B2M*-aligned reads using the Integrative Genomics Viewer (IGV)[48]. Here mutations were phased in case single reads were observed spanning the genomic locations of both mutations. We divided patients with multiple *B2M* mutations into three subgroups: (1) Biallelic, if (a) the multiple mutations were in *trans* AND the integer sum of the mutation copy numbers equalled (or exceeded) the integer copy number of the *B2M* gene (for mutations in *cis*, only one of these mutations was considered in the calculation); or (b) at least one of the mutations showed LOH. (2) Potentially biallelic, if the multiple mutations affected genomic locations too distant to be phased and the (integer) sum of the mutation copy numbers equalled (or exceeded) the integer copy number of the *B2M* gene (for mutations in *cis*, only one of these mutations was considered in the calculation) and none of the mutations showed LOH. (3) Not bi-allelic, if the integer sum of the mutation copy numbers was smaller than the integer copy number of the *B2M* gene (for mutations in *cis*, only one of these mutations was considered in the calculation) and none of the mutations showed LOH.

In these analyses, mutations were considered to be subclonal in the case in which the probability of subclonality was >0.5 (the situation in which a mutation is more likely subclonal than clonal), as determined using PURPLE (v.2.34)[47].

Association of *B2M* status with outcome and tumour characteristics

To test whether somatic *B2M* alterations were associated with the clinical benefit rate of patients with MMR-d tumours treated with ICB in the DRUP, we used a Fisher's exact test (using the Python package Scipy[49] (v.1.3.1)) for unadjusted analyses and logistic regression (as implemented by the Python package Statsmodels (https://pypi.org/project/statsmodels/; v.0.10.1) for analyses adjusted for the continuous TMB per Mb and/or the primary site of the tumour. The association of *B2M* status with TMB was tested using Scipy's Wilcoxon rank-sum test. Associations of *B2M* status with the primary site of the tumour or the biopsy location were tested using Scipy's Fisher's exact test.

Association of TMB with ICB treatment outcome

For the DRUP cohort, the association of clinical benefit with TMB was tested using Statsmodels' Wilcoxon rank-sum test.

Differential gene expression analysis

Differential RNA expression of genes was tested in R using EdgeR[50] (v.3.28.1) and Limma[51] (including Voom[52]) (v.3.42.2). Raw read counts were filtered by removing low-expressed genes. Normalization factors were calculated using EdgeR to transform the raw counts to $\log_2$[counts per million reads (CPM)] and calculate residuals using Voom. Voom was then used to fit a smoothened curve to the √(residual standard deviation) by average gene expression, which was then plotted for visual inspection to confirm that the appropriate threshold was used for filtering of low-expressed genes (defined as the minimal amount of filtering necessary to overcome a dipping mean-variance trend at low counts). Next, Limma was used to calculate the differential expression of genes on the basis of a linear model fit, considering the smoothened curve for sample weights, and empirical Bayes smoothing of standard errors. FDR-adjusted *P* values were calculated using Benjamini–Hochberg correction of the obtained *P* values.

TCGA

Using TCGA data, we calculated the differential expression between tumours with and without mutations in *B2M*, adjusting for tumour type, using the following design formula: expression ∝ Primary_Site + B2M_status (+ intercept by default), for which Primary_Site was a three-levelled factor (COAD, STAD, or UCEC) and B2M_status was a two-levelled factor (mutated, or wildtype).

NICHE study

Using NICHE study data, we calculated differential expression between pre- and post-ICB treatment. To respect the paired nature of these data, we used the following design formula: expression ∝ Patient + ICB + intercept, where Patient is a factor for each individual patient and ICB is a two-levelled factor (ICB-treated, yes/no).

Immune marker gene set expression analysis

To use RNA-seq data to obtain a relative estimate of the infiltration of specific immune cell types within tumours, we calculated the average $\log_2[\text{RPM} + 1]$ expression of marker genes that are specifically expressed in the immune cell types of interest. To this end, we used the previously published marker gene sets[24], and extended this by (1) *CD4* as a $CD4^+$ T cell marker gene; (2) *TRDV1* and *TRDV3* as γδ1/3T cell marker genes; and (3) a killer-cell Ig-like receptor (KIR) gene set (comprising all genes of which the name starts with KIR and of which the name contains DL or DS). We excluded the gene set 'NK $CD56^{dim}$ cells' of ref. [24] (comprising *IL21R*, *KIR2DL3*, *KIR3DL1* and *KIR3DL2*) from our analyses, as three out of four genes within this set were KIRs and this set therefore showed high collinearity/redundancy to our full KIR gene set. As *XLC1* and *XLC2* are highly expressed by tumour-infiltrating γδ T cells, these genes were removed from the NK cell marker gene set and replaced by *KLRF1*, which encodes the well-established NK cell marker NKp80. The resulting gene set consisted of *NCR1* and *KLRF1*, encoding the well-established NK cell markers NKp46 and NKp80, respectively. Finally, we reduced the 'cytotoxic cells' marker gene set of ref. [24] to those genes in the set encoding cytotoxic

molecules (*GZMA*, *GZMB*, *GZMH*, *PRF1*, *GNLY*, *CTSW*). A list of the final collection of our marker gene sets is provided in Supplementary Table 2.

Association of immune cell marker gene set expression with *B2M* alteration status (alteration yes/no) was calculated as follows:

1. 1.

   For TCGA-study-based analyses, we used (i) the Wilcoxon rank-sum test (for unadjusted analyses) and (ii) ordinary least squares linear regression (for analyses adjusted for tumour type; as implemented by the Python package ‘Statsmodels), using a similar design formula as for the differential gene expression analysis.

2. 2.

   For DRUP-cohort-based analyses, we used a linear mixed effects model (as implemented by the lmer function of the R package Lme4 (v.1.1.26)), adjusting for tumour type and biopsy site as random effects, using the following design formula: expression ∝ B2M_status + (1|tumour_type) + (1|biopsy_site) + intercept. In subgroup analysis of CRC, we omitted tumour type in this formula: expression ∝ B2M_status + (1|biopsy_site) + intercept.

3. 3.

   For Hartwig data-based analyses, we used ordinary least squares linear regression (as implemented by the Python package Statsmodels) adjusting for tumour type, using the following design formula: expression ∝ Primary_Site + B2M_status + intercept.

4. 4.

   For NICHE-study-based analyses, we used the Wilcoxon rank-sum test (as implemented by the Python package Scipy).

Hierarchical clustering

Hierarchical clustering of expression profiles of individual genes or immune marker gene sets of TCGA cohorts was performed on Z-score-transformed $\log_2$[RPM + 1] expression values, using the Python package Scipy[49], with Euclidean distance as the distance metric and using the Ward variance minimization algorithm. Here, we used the default settings with one exception—for visualization purposes, the colour threshold was halved in the TCGA-based clustering of individual genes.

Patient samples

The DRUP study and the generation of the Hartwig database were initiated and conducted on behalf of the Center for Personalized Cancer Treatment (CPCT; ClinicalTrials.gov: NCT02925234, NCT01855477). These studies were approved by the Medical Ethical Committee of the Netherlands Cancer Institute in Amsterdam and the University Medical Center Utrecht, respectively, and were conducted in accordance with good clinical practice guidelines and the Declaration of Helsinki's ethical principles for medical research. Written informed consent was obtained from all of the study participants. Moreover, primary colon cancer tissues from a total of 17 patients with colon cancer who underwent surgical resection of their tumour at the Leiden University Medical Center (LUMC, The Netherlands; Supplementary Table 4) were used for scRNA-seq, IMC and functional assays. No patient with a previous history of inflammatory bowel disease was included. This study was approved by the Medical Ethical Committee of the Leiden University Medical Center (protocol P15.282), and patients provided written informed consent. Finally, primary colon cancer tissues from ten patients with colon cancer included in the NICHE study (NCT03026140)[9] carried out at the Netherlands Cancer Institute (NKI, The Netherlands) were used for this study. All samples were anonymized and handled according to the ethical guidelines described in the Code for Proper Secondary Use of Human Tissue in the Netherlands of the Dutch Federation of Medical Scientific Societies.

Processing of colon cancer tissues

Details on the processing of colon cancer tissues have been described previously[22]. In brief, macroscopic sectioning from the lumen to the most

invasive area of the tumour was performed. Tissues were collected in IMDM + GlutaMax medium (Gibco) complemented with 20% fetal calf serum (FCS) (Sigma-Aldrich), 1% penicillin–streptomycin (Gibco) and fungizone (Gibco), and 0.1% ciprofloxacin (provided by apothecary LUMC) and gentamicin (Invitrogen), and immediately cut into small fragments in a Petri dish. Enzymatic digestion was performed using 1 mg ml^{-1} collagenase D (Roche Diagnostics) and 50 μg ml^{-1} DNase I (Roche Diagnostics) in 5 ml of IMDM + GlutaMax medium for 30 min at 37 °C in gentleMACS C tubes (Miltenyi Biotec). During and after incubation, cell suspensions were dissociated mechanically on the gentleMACS Dissociator (Miltenyi Biotec). Cell suspensions were filtered through a 70 μm cell strainer (Corning), washed in IMDM + GlutaMax medium with 20% FCS, 1% penicillin–streptomycin and 0.1% fungizone, and the cell count and viability were determined using the Muse Count & Viability Kit (Merck) on the Muse Cell Analyser (Merck). On the basis of the number of viable cells, cells in IMDM + GlutaMax medium were cryopreserved in liquid nitrogen until time of analysis complemented 1:1 with 80% FCS and 20% dimethyl sulfoxide (Merck).

Immunohistochemical detection of MMR, B2M and HLA class I proteins

For the tumour tissue samples from the LUMC, tumour MMR status was determined by immunohistochemical detection of PMS2 (anti-PMS2 antibodies; EP51, DAKO) and MSH6 (anti-MSH6 antibodies; EPR3945, Abcam) proteins[53]. MMR-deficiency was defined as the lack of expression of at least one of the MMR-proteins in the presence of an internal positive control. Tumour B2M status was determined by immunohistochemical detection of B2M (anti-B2M antibodies; EP2978Y, Abcam). Immunohistochemical detection of HLA class I expression on tumour cells was performed with HCA2 and HC10 monoclonal antibodies (Nordic-MUbio), and tumours were classified as HLA class I positive, weak or loss, as described previously[16]. For the tumour samples from the NICHE study, immunohistochemistry analysis of the formalin-fixed paraffin-embedded (FFPE) tissue was performed on the BenchMark Ultra autostainer (Ventana Medical Systems). In brief, paraffin sections were cut at 3 μm, heated at 75 °C for 28 min and deparaffinized in the instrument using EZ prep solution

(Ventana Medical Systems). Heat-induced antigen retrieval was performed using Cell Conditioning 1 (CC1, Ventana Medical Systems) for 32 min at 95 °C (HC10) or 64 min at 95 °C (B2M and HCA2). HLA class I heavy chain expression was detected using clone HCA2 (1:5,000, 60 min at room temperature; Nordic-Mubio) and clone HC10 (1:20,000, 32 min at 37 °C; Nordic-Mubio). B2M was detected using clone D8P1H (1:1,500, 60 min at room temperature; Cell Signaling). Bound antibodies were detected using the OptiView DAB Detection Kit (Ventana Medical Systems). Slides were counterstained with haematoxylin and Bluing Reagent (Ventana Medical Systems). A PANNORAMIC 1000 scanner from 3DHISTECH was used to scan the slides at ×40 magnification.

Sorting of γδ T cells from colon cancers and scRNA-seq

scRNA-seq was performed on sorted γδ T cells from colon cancers (MMR-d) of five patients from the LUMC in the presence of hashtag oligos (HTOs) for sample ID and antibody-derived tags (ADTs) for CD45RA and CD45RO protein expression by CITE-seq[54]. Cells were thawed, rested at 37 °C in IMDM (Lonza)/20% FCS for 1 h, and then incubated with human Fc receptor block (BioLegend) for 10 min at 4 °C. Cells were then stained with cell surface antibodies (anti-CD3-PE (1:50, SK7, BD Biosciences), anti-CD45-PerCP-Cy5.5 (1:160, 2D1, eBioscience), anti-CD7-APC (1:200, 124-1D1, eBioscience), anti-EPCAM-FITC (1:60, HEA-125, Miltenyi), anti-TCRγδ-BV421 (1:80, 11F2, BD Biosciences); and a 1:1,000 near-infrared viability dye (Life Technologies)), 1 µg of TotalSeq-C anti-CD45RA (HI100, BioLegend, 2 µl per sample) and 1 µg of TotalSeq-C anti-CD45RO (UCHL1, BioLegend, 2 µl per sample) antibodies, and 0.5 µg of a unique TotalSeq-C CD298/B2M hashtag antibody (LNH-94/2M2, BioLegend, 1 µl per sample) for 30 min at 4 °C. Cells were washed three times in FACS buffer (PBS (Fresenius Kabi)/1% FCS) and kept under cold and dark conditions until cell sorting. Compensation was performed using CompBeads (BD Biosciences) and ArC reactive beads (Life Technologies). Single, live $CD45^{+}EPCAM^{-}CD3^{+}TCR\gamma\delta^{+}$ cells were sorted on the FACS Aria III 4L (BD Biosciences) system. After sorting, the samples were pooled.

scRNA-seq libraries were prepared using the Chromium Single Cell 5′ Reagent Kit v1 chemistry (10x Genomics) according to the manufacturer's instructions. The construction of 5′ gene expression libraries enabled the identification of γδ T cell subsets according to Vδ and Vγ usage. Libraries were sequenced on the HiSeq X Ten using paired-end 2 × 150 bp sequencing (Illumina). Reads were aligned to the human reference genome (GRCh38) and quantified using Cell Ranger (v.3.1.0). Downstream analysis was performed using Seurat (v.3.1.5) according to the author's instructions[55]. In brief, cells that had less than 200 detected genes and genes that were expressed in less than six cells were excluded. The resulting 5,669 cells were demultiplexed on the basis of HTO enrichment using the MULTIseqDemux algorithm[56]. Next, cells with a mitochondrial gene content of greater than 10% and cells with outlying numbers of expressed genes (>3,000) were filtered out from the analysis, resulting in a final dataset of 4,442 cells, derived from HTO1 ($n = 332$), HTO6 ($n = 105$), HTO7 ($n = 1,100$), HTO8 ($n = 1,842$) and HTO9 ($n = 1,063$). Data were normalized using the LogNormalize function of Seurat with a scale factor of 10,000. Variable features were identified using the FindVariableFeatures function of Seurat returning 2,000 features. We next applied the RunFastMNN function of SeuratWrappers split by sample ID to adjust for potential batch-derived effects across the samples[57]. Uniform manifold approximation and projection (UMAP)[58] was used to visualize the cells in a two-dimensional space, followed by the FindNeighbors and FindClusters functions of Seurat. Data were scaled, and heterogeneity associated with mitochondrial contamination was regressed out. Cell clusters were identified by performing differentially expressed gene analysis using the FindAllMarkers function, with min.pct and logfc.threshold at 0.25. The number of $TRDV1^+$ (Vδ1, $n = 1,927$), $TRDV2^+$ (Vδ2, $n = 860$) or $TRDV3^+$ (Vδ3, $n = 506$) cells was determined as the percentage of all cells with an expression level of >1, with <1 for the other TCR Vδ chains. CRC96, 134 and 167 had less than ten $TRDV3^+$ cells, and were not included in the Vδ3 analysis. Transcripts of Vδ4 (*TRDV4*), Vδ5 (*TRDV5*) and Vδ8 (*TRDV8*) cells were not detected. The percentage of cells positive for a certain gene was determined as all cells with an expression level of >1.

IMC staining and analysis

IMC analysis was performed on ICB-naive colon cancer tissues (MMR-d) of 17 patients from the LUMC; 5 of these colon cancer tissues had B2M defects and the remainder were B2M-positive (Supplementary Table 4). Moreover, IMC was performed on ICB-naive and ICB-treated colon cancer tissues (MMR-d) of ten patients from the NICHE study; five of these colon cancer tissues were $B2M^{WT}$ and five were $B2M^{MUT}$. Antibody conjugation and immunodetection were performed according to previously published methodology[59]. FFPE tissue (thickness, 4 μm) was incubated with 41 antibodies in four steps. First, sections were incubated overnight at room temperature with anti-CD4 and anti-TCRδ antibodies, which were subsequently detected using metal-conjugated secondary antibodies (1 μg ml^{-1}, donkey anti-rabbit IgG and goat anti-mouse IgG, respectively; Abcam). Second, the sections were incubated with 20 antibodies (Supplementary Table 6) for five hours at room temperature. Third, the sections were incubated overnight at 4 °C with the remaining 19 antibodies (Supplementary Table 6). Fourth, the sections were incubated with 0.125 μM Cell-ID intercalator-Ir (Fluidigm) to detect the DNA, and stored dry until measurement. For each sample, six 1,000 μm × 1,000 μm regions (two to three for pretreatment NICHE biopsies due to the small tissue size) were selected on the basis of consecutive haematoxylin and eosin stains and ablated using the Hyperion Imaging system (Fluidigm). Data were acquired using the CyTOF Software (v.7.0) and exported using MCD Viewer (v.1.0.5). Data were normalized using semi-automated background removal in ilastik[60] (v.1.3.3), to control for variations in the signal-to-noise ratio between FFPE sections as described previously[61]. Next, the phenotype data were normalized at the pixel level. Cell segmentation masks were created for all cells in ilastik and CellProfiler[62] (v.2.2.0). In ImaCytE[63] (v.1.1.4), cell segmentation masks and normalized images were combined to generate single-cell FCS files containing the relative frequency of positive pixels for each marker per cell. Cells forming visual neighbourhoods in a *t*-distributed stochastic neighbour embedding[64] in Cytosplore[65] (v.2.3.0) were grouped and exported as separate FCS files. The resulting subsets were imported back into ImaCyte and visualized on the segmentation masks. Expression of immunomodulatory markers was determined as all cells with a relative frequency of at least 0.2 positive pixels per cell. Differences in cells per mm^2 were calculated using Mann–Whitney tests in GraphPad Prism

(v.9.0.1). Image acquisition and analysis were performed blinded to group allocation.

Sorting of γδ T cells from colon cancers and cell culturing

γδ T cells from colon cancers (MMR-d) of five patients from the LUMC were sorted for cell culture. Cells were thawed and rested at 37 °C in IMDM (Lonza)/10% nHS for 1 h. Next, cells were incubated with human Fc receptor block (BioLegend) and stained with cell surface antibodies (anti-CD3-Am Cyan (1:20, SK7, BD Biosciences), anti-TCRγδ-BV421 (1:80, 11F2, BD Biosciences) and anti-PD-1-PE (1:30, MIH4, eBioscience)) for 45 min at 4 °C together with different additional antibodies for immunophenotyping (anti-CD103-FITC (1:10, Ber-ACT8, BD Biosciences), anti-CD38-PE-Cy7 (1:200, HIT2, eBioscience); anti-CD39-APC (1:60, A1, BioLegend), anti-CD45RA-PE-Dazzle594 (1:20, HI100, Sony), anti-CD45RO-PerCP-Cy5.5 (1:20, UCHL1, Sony), anti-TCRαβ-PE-Cy7 (1:40, IP26, BioLegend), anti-TCRVδ1-FITC (1:50, TS8.2, Invitrogen) or anti-TCRVδ2-PerCP-Cy5.5 (1:200‘ B6, BioLegend). A 1:1,000 live/dead fixable near-infrared viability dye (Life Technologies) was included in each staining. Cells were washed three times in FACS buffer (PBS/1% FCS) and kept under cold and dark conditions until cell sorting. Compensation was performed using CompBeads (BD Biosciences) and ArC reactive beads (Life Technologies). Single, live $CD3^{+}TCR\gamma\delta^{+}PD\text{-}1^{+}$ and $PD\text{-}1^{-}$ cells were sorted on the FACS Aria III 4L (BD Biosciences) system. For CRC94, all γδ T cells were sorted owing to the low number of $PD\text{-}1^{+}$ cells. γδ T cells were sorted in medium containing feeder cells (1×10^{6} per ml), PHA (1 μg ml^{-1}; Thermo Fisher Scientific), IL-2 (1,000 IU ml^{-1}; Novartis), IL-15 (10 ng ml^{-1}; R&D Systems), gentamicin (50 μg ml^{-1}) and fungizone (0.5 μg ml^{-1}). Sorted γδ T cells were expanded in the presence of 1,000 IU ml^{-1} IL-2 and 10 ng ml^{-1} IL-15 for 3–4 weeks. The purity and phenotype of γδ T cells were assessed using flow cytometry. We obtained a >170,000-fold increase in 3–4 weeks of expansion of γδ T cells (Extended Data Fig. 4c).

Immunophenotyping of expanded γδ T cells by flow cytometry

Expanded γδ T cells from colon tumours were analysed by flow cytometry for the expression of TCR Vδ chains, NKG2 receptors, NCRs, KIRs, tissue-

residency/activation markers, cytotoxic molecules, immune checkpoint molecules, cytokine receptors and Fc receptors. In brief, cells were incubated with human Fc receptor block (BioLegend) and stained with cell surface antibodies (Supplementary Table 7) for 45 min at 4 °C, followed by three wash steps in FACS buffer (PBS/1% FCS). Granzyme B and perforin were detected intracellularly using fixation buffer and intracellular staining permeabilization wash buffer (BioLegend). Compensation was performed using CompBeads (BD Biosciences) and ArC reactive beads (Life Technologies). Cells were acquired on the FACS LSR Fortessa 4L (BD Biosciences) system running FACSDiva software (v.9.0; BD Biosciences). Data were analysed using FlowJo (v.10.6.1; Tree Star).

Cancer cell line models and culture

Human colorectal adenocarcinoma cell lines HCT-15 (MMR-d), LoVo (MMR-d), HT-29 (MMR-p), SW403 (MMR-p) and SK-CO-1 (MMR-p), as well as HLA-class-I-deficient human leukaemia cell line K-562 and Burkitt lymphoma cell line Daudi were used as targets for reactivity and immune cell killing assays. All of the cell lines were obtained from the ATCC. The cell lines were authenticated by STR profiling and tested for mycoplasma. HCT-15, LoVo, HT-29, K-562 and Daudi cells were maintained in RPMI (Gibco)/10% FCS. SW403 and SK-CO-1 were maintained in DMEM/F12 (Gibco)/10% FCS. All adherent cell lines were trypsinized before passaging. The *B2M*-knockin HCT-15 and LoVo cell lines were generated using the *B2M* plasmid (pLV[Exp]-EF1A>hB2M[NM_004048.4](ns):T2A:Puro), produced in lentivirus according to standard methodology. Cells were selected using puromycin and then FACS-sorted on the basis of HLA-A/B/C expression using 1:100 anti-HLA-A/B/C-FITC (W6/32, eBioscience).

Organoid models and culture

Tumour organoids were derived from MMR-d CRC tumours of two patients through resection from the colon (tumour organoid 1) or peritoneal biopsy (tumour organoid 2) (Supplementary Table 5). Establishment of the respective organoid lines from tumour material was performed as previously reported[66,67]. In brief, tumour tissue was mechanically dissociated and digested with 1.5 mg ml^{-1} of collagenase II (Sigma-Aldrich), 10 μg ml^{-1} of

hyaluronidase type IV (Sigma-Aldrich) and 10 μM Y-27632 (Sigma-Aldrich). Cells were embedded in Cultrex RGF BME type 2 (3533-005-02, R&D systems) and placed into a 37 °C incubator for 20 min. Human CRC organoid medium is composed of Ad-DF+++ (Advanced DMEM/F12 (GIBCO) supplemented with 2 mM Ultraglutamine I (Lonza), 10 mM HEPES (GIBCO), 100 U ml^{-1} of each penicillin and streptomycin (GIBCO), 10% noggin-conditioned medium, 20% R-spondin1-conditioned medium, 1× B27 supplement without vitamin A (GIBCO), 1.25 mM *N*-acetylcysteine (Sigma-Aldrich), 10 mM nicotinamide (Sigma-Aldrich), 50 ng ml^{-1} human recombinant EGF (Peprotech), 500 nM A83-01 (Tocris), 3 μM SB202190 (Cayman Chemicals) and 10 nM prostaglandin E2 (Cayman Chemicals). Organoids were passaged depending on growth every 1–2 weeks by incubating in TrypLE Express (Gibco) for 5–10 min followed by embedding in BME. Organoids were authenticated by SNP array or STR analysis and were regularly tested for Mycoplasma using Mycoplasma PCR43 and the MycoAlert Mycoplasma Detection Kit (LT07-318). In the first two weeks of organoid culture, 1× Primocin (Invivogen) was added to prevent microbial contamination. Procedures performed with patient samples were approved by the Medical Ethical Committee of the Netherlands Cancer Institute–Antoni van Leeuwenhoek hospital (NL48824.031.14) and written informed consent was obtained from all of the patients. Mismatch repair status was assessed using a standard protocol for the Ventana automated immunostainer for MLH1 clone M1 (Roche), MSH2 clone G219-1129 (Roche), MSH6 clone EP49 (Abcam) and PMS2 clone EP51 (Agilant Technologies). The *B2M*^{KO} tumour organoid lines were generated using sgRNA targeting *B2M* (GGCCGAGATGTCTCGCTCCG), cloned into LentiCRISPR v2 plasmid. The virus was produced using a standard method.

Screening of cancer cell lines and tumour organoids by flow cytometry

The cancer cell lines used in the reactivity and killing assays were screened for the expression of B2M, HLA class I molecules, NKG2D ligands, DNAM-1 ligands and butyrophilin using flow cytometry. In brief, cells were incubated with human Fc receptor block (BioLegend) and stained with the cell surface antibodies in different experiments (anti-CD112-PE (1:10, R2.525, BD Biosciences), anti-CD155-PE (1:10, 300907, R&D Systems),

anti-CD277/BTN3A1-PE (1:50, BT3.1, Miltenyi), anti-B2M-PE (1:100, 2M2, BioLegend), anti-HLA-A/B/C-FITC (1:100, W6/32, eBioscience), anti-HLA-A/B/C-AF647 (1:160, W6/32, BioLegend), anti-HLA-E-BV421 (1:20, 3D12, BioLegend), anti-HLA-G-APC (1:20, 87G, BioLegend), anti-MICA/B-PE (1:300, 6D4, BioLegend), anti-ULBP1-PE (1:10, 170818, R&D Systems), anti-ULBP2/5/6-PE (1:20, 165903, R&D Systems), anti-ULBP3-PE (1:20, 166510, R&D Systems) or anti-ULBP4-PE (1:20, 709116, R&D Systems)) for 45 min at 4 °C. A 1:1,000 live/dead fixable near-infrared viability dye (Life Technologies) was included in each staining. Cells were washed three times in FACS buffer (PBS/1% FCS). Compensation was performed using CompBeads (BD Biosciences) and ArC reactive beads (Life Technologies). Cells were acquired on the FACS Canto II 3L or FACS LSR Fortessa 4L (BD Biosciences) system running FACSDiva software (v.9.0; BD Biosciences). Isotype or FMO controls were included to determine the percentage of positive cancer cells. Data were analysed using FlowJo v.10.6.1 (Tree Star).

For organoid surface staining, tumour organoids were dissociated into single cells using TrypLE Express (Gibco), washed twice in cold FACS buffer (PBS, 5 mM EDTA, 1% bovine serum antigen), and stained with anti-HLA-A/B/C-PE (1:20, W6/32, BD Biosciences), anti-B2M-FITC (1:100, 2M2, BioLegend), anti-PD-L1-APC (1:200, MIH1, eBioscience) and 1:2,000 near-infrared (NIR) viability dye (Life Technologies), or isotype controls (1:1,000 FITC; 1:20, PE; or 1:200, APC) mouse IgG1 kappa (BD Biosciences). For NKG2D ligand expression analysis, cells were stained with anti-MICA/MICB (1:300), anti-ULBP1 (1:10), anti-ULBP2/5/6 (1:20), anti-ULBP3 (1:20), anti-ULBP4 (1:20) and 1:2,000 near-infrared (NIR) viability dye (Life Technologies). Tumour cells were incubated for 30 min at 4 °C in the dark and washed twice with FACS buffer. All of the samples were recorded with the BD LSR Fortessa Cell Analyzer SORP flow cytometer using FACSDiVa (v.8.0.2; BD Biosciences). Data were analysed using FlowJo (v.10.6.1; BD) and presented using GraphPad Prism (v.9.0.0; GraphPad).

Reactivity assay of γδ T cells

The reactivity of γδ T cells to the different cancer cell lines was assessed by a co-culture reactivity assay. γδ T cells were thawed and cultured in IMDM + GlutaMax (Gibco)/8% nHS medium with penicillin (100 IU ml^{-1}) and streptomycin (100 μg ml^{-1}) in the presence of low-dose IL-2 (25 IU ml^{-1}) and IL-15 (5 ng ml^{-1}) overnight at 37 °C. Cancer cell lines were counted, adjusted to a concentration of 0.5×10^5 cells per ml in IMDM + GlutaMax/10% FCS medium with penicillin (100 IU ml^{-1}) and streptomycin (100 μg ml^{-1}), and seeded (100 μl per well) in coated 96-well flat-bottom microplates (Greiner CellStar) (for 5,000 cells per well) overnight at 37 °C. The next day, γδ T cells were collected, counted and adjusted to a concentration of 1.2×10^6 cells per ml in IMDM + GlutaMax/10% FCS medium. The γδ T cells were added in 50 μl (for 60,000 cells per well) and co-cultured (12:1 effector:target ratio) at 37 °C for 18 h in biological triplicates. The medium (without cancer cells) was used as a negative control and PMA (20 ng ml^{-1})/ionomycin (1 μg ml^{-1}) was used as a positive control. After co-culture, the supernatant was collected to detect IFNγ secretion by enzyme-linked immunosorbent assay (Mabtech) according to the manufacturer’s instructions. Moreover, cells were collected, incubated with human Fc receptor block (BioLegend) and stained with cell surface antibodies (anti-CD137-APC (1:100, 4B4-1, BD Biosciences), anti-CD226/DNAM-1-BV510 (1:150, DX11, BD Biosciences), anti-CD3-AF700 (1:400, UCHT1, BD Biosciences), anti-CD39-APC (1:80, A1, BioLegend), anti-CD40L-PE (1:10, TRAP1, BD Biosciences), or anti-PD-1-PE (1:30, MIH4, eBioscience), anti-TCRγδ-BV650 (1:40, 11F2, BD Biosciences), anti-NKG2D-PE-Cy7 (1:300, 1D11, BD Biosciences) and anti-OX40-FITC (1:20, ACT35, BioLegend)) for 45 min at 4 °C. A 1:1,000 live/dead fixable near-infrared viability dye (Life Technologies) was included in each staining. Cells were washed three times in FACS buffer (PBS/1% FCS). Compensation was performed using CompBeads (BD Biosciences) and ArC reactive beads (Life Technologies). Cells were acquired on the FACS LSR Fortessa X-20 4L (BD Biosciences) system running FACSDiva software (v.9.0; BD Biosciences). Data were analysed using FlowJo (v.10.6.1; Tree Star). All data are representative of at least two independent experiments.

Immune cell killing assay γδ T cells

Killing of the different cancer cell lines by γδ T cells was visualized and quantified by a co-culture immune cell killing assay using the IncuCyte S3 Live-Cell Analysis System (Essen Bioscience). HCT-15, LoVo and HT-29 cells were transduced with IncuCyte NucLight Red Lentivirus Reagent (EF-1α, Puro; Essen BioScience) providing a nuclear-restricted expression of a red (mKate2) fluorescent protein. In brief, HCT-15, LoVo and HT-29 cells were seeded, transduced according to the manufacturer's instructions and stable cell populations were generated using puromycin selection. The *B2M*-knockin cell lines were created under puromycin selection; therefore, stable NucLight Red-expressing cell populations were generated by sorting for mKate2 (the red fluorescent protein) in the PE Texas Red filter set instead. Cancer cell lines were counted, adjusted to a concentration of 1×10^5 cells per ml in IMDM + GlutaMax/10% FCS medium with penicillin (100 IU ml^{-1}) and streptomycin (100 μg ml^{-1}), and seeded (100 μl per well) in 96-well flat-bottom clear microplates (Greiner CellStar) (for 10,000 cells per well). The target cell plate was placed in the IncuCyte system at 37 °C to monitor for cell confluency for 3 days. On day 2, γδ T cells were thawed and cultured in IMDM + GlutaMax/8% nHS medium with penicillin (100 IU ml^{-1}) and streptomycin (100 μg ml^{-1}) in the presence of low-dose IL-2 (25 IU ml^{-1}) and IL-15 (5 ng ml^{-1}) overnight at 37 °C. The next day, γδ T cells were collected, counted and adjusted to a concentration of 7.2×10^5 cells per ml in IMDM + GlutaMax/10% FCS medium. After aspiration of the medium of the target cell plate, 100 μl of new medium containing 3.75 μM IncuCyte Caspase-3/7 Green Apoptosis Reagent (Essen BioScience) (1.5× final assay concentration of 2.5 μM) was added together with 50 μl of γδ T cells (for 36,000 cells per well). They were co-cultured (4:1 effector:target ratio) in the IncuCyte system at 37 °C in biological duplicates. Cancer cells alone and cancer cells alone with caspase-3/7 were used as negative controls. Images (2 images per well) were captured every hour at ×20 magnification with the phase, green and red channels for up to 4 days.

Analysis was performed using the IncuCyte software (v.2020B) for each cancer cell line separately. The following analysis definitions were applied: a minimum phase area of 200 μm^2, RCU of 2.0, and a GCU of 2.0 (for HCT-15 cells) and 4.0 (for LoVo and HT-29 cells). Cancer cell apoptosis was then quantified in the IncuCyte software by counting the total number of green + red objects per image normalized (by division) to the total number of red

objects per image after 12 h co-culture and displayed as a percentage (mean ± s.e.m.) of two wells with two images per well. For the comparison of the killing of *B2M*-knockin HCT-15 and LoVo cell lines versus the WT cell lines, Caspase-3/7 Red Apoptosis Reagent (Essen BioScience) was used. The transfection of the target reporter was not as successful in combination with the *B2M*-knockin. Thus, apoptosis was quantified by dividing the red area by the phase area and displayed as a percentage (mean ± s.e.m.) of two wells with two images per well. The following analysis definitions were applied: a minimum phase area of 100 μm^2 and a RCU of 0.5 (for HCT-15 cells) and 0.75 (for LoVo cells).

Tumour organoid recognition assay

For evaluation of tumour reactivity towards $B2M^{WT}$ and $B2M^{KO}$ organoids and NKG2D ligand blocking conditions, tumour organoids and γδ T cells were prepared as described previously[9,66,67]. Two days before the experiment, organoids were isolated from BME by incubation in 2 mg ml^{-1} type II dispase (Sigma-Aldrich) for 15 min before addition of 5 mM EDTA and washed with PBS before being resuspended in CRC organoid medium with 10 μM Y-27632 (Sigma-Aldrich). The organoids were stimulated with 200 ng ml^{-1} IFNγ (Peprotech) 24 h before the experiment. For the recognition assay and intracellular staining, tumour organoids were dissociated into single cells and plated in anti-CD28-coated (CD28.2, eBioscience) 96-well U-bottom plates with γδ T cells at a 1:1 target:effector ratio in the presence of 20 μg ml^{-1} anti-PD-1 (Merus). As a positive control, γδ T cells were stimulated with 50 ng ml^{-1} of phorbol-12-myristate-13-acetate (Sigma-Aldrich) and 1 μg ml^{-1} of ionomycin (Sigma-Aldrich). After 1 h of incubation at 37 °C, GolgiSTOP (BD Biosciences, 1:1,500) and GolgiPlug (BD Biosciences, 1:1,000) were added. After 4 h of incubation at 37 °C, γδ T cells were washed twice in cold FACS buffer (PBS, 5 mM EDTA, 1% bovine serum antigen) and stained with anti-CD3-PerCP-Cy5.5 (1:20, BD Biosciences), anti-TCRγδ-PE (1:20, BD Bioscience), anti-CD4-FITC (1:20, BD Bioscience) (not added in experiments with NKG2D ligand blocking), anti-CD8-BV421 (1:200, BD Biosciences) and 1:2,000 near-infrared (NIR) viability dye (Life Technologies) for 30 min at 4 °C. Cells were washed, fixed and stained with 1:40 anti-IFNγ-APC (BD Biosciences) for 30 min at 4 °C, using the Cytofix/Cytoperm Kit (BD Biosciences). After

two wash steps, cells were resuspended in FACS buffer and recorded with the BD LSR Fortessa Cell Analyzer SORP flow cytometer using FACSDiVa software (v.8.0.2; BD Biosciences). Data were analysed using FlowJo (v.10.6.1, BD) and presented using GraphPad Prism (v.9.0.0, GraphPad).

Blocking experiments with cancer cell lines and tumour organoids

The reactivity of and killing by the γδ T cells was examined in the presence of different blocking antibodies to investigate which receptor–ligand interactions were involved. For DNAM-1 blocking, γδ T cells were incubated with 3 μg ml^{-1} purified anti-DNAM-1 (DX11, BD Biosciences) for 1 h at 37 °C. For γδ TCR blocking, γδ T cells were incubated with 3 μg ml^{-1} purified anti-TCRγδ (5A6.E9, Invitrogen) for 1 h at 37 °C; the clone that we used was tested to be the best for use in γδ TCR blocking assays[68]. NKG2D ligands were blocked on the cancer cell lines and single cells of tumour organoids by incubating the target cells with 12 μg ml^{-1} anti-MICA/B (6D4, BioLegend), 1 μg ml^{-1} anti-ULBP1 (170818, R&D Systems), 3 μg ml^{-1} anti-ULBP2/5/6 (165903, R&D Systems) and 6 μg ml^{-1} anti-ULBP3 (166510, R&D Systems) for 1 h at 37 °C before plating with γδ T cells. After incubation with the blocking antibodies, the γδ T cells were added to cancer cell lines HCT-15, LoVo and HT-29 as described above with a minimum of two biological replicates per blocking condition. For organoid experiments, anti-CD107a-FITC (1:50, H4A3, BioLegend) was added during incubation.

As a control for Fc-mediated antibody effector functions, γδ T cells alone were incubated with the blocking antibodies in the presence of 2.5 μM IncuCyte Caspase-3/7 Green Apoptosis Reagent (Essen BioScience) in the IncuCyte system at 37 °C, and the number of apoptotic γδ T cells was quantified over time.

Data analysis and visualization

Bulk DNA-seq and RNA-seq data were analysed using Python (v.3) and R (v.3.6.1) in Jupyter Notebook (v.6.0.1). Numpy (v.1.17.2) and Pandas (v.0.25.1) were used for array and data frame operations, respectively. Data

visualization was performed using Matplotlib (v.3.2.1) and Seaborn (v.0.9.0). scRNA-seq data were analysed using Cell Ranger (v.3.1.0), R (v.4.1.0) and Seurat (v.3.1.5). IMC data were analysed using ilastik (v.1.3.3), CellProfiler (v.2.2.0), ImaCytE (v.1.1.4) and Cytosplore (v.2.3.0). Flow cytometry data were analysed using FlowJo (v.10.6.1). IncuCyte data were analysed using IncuCyte (v.2020B). Data visualization was performed using GraphPad Prism (v.9.0.0 and v.9.0.1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The TCGA data used here are publicly available at the National Cancer Institute GDC Data Portal (https://portal.gdc.cancer.gov; cohorts COAD, STAD and UCEC). Of the DRUP study participants included in this preliminary analysis across all (complete and incomplete) cohorts of the study, we included all clinical data, genomics data on *B2M* status and RNA expression data of marker gene sets in Supplementary Table 1. The raw sequencing data of the DRUP and Hartwig cohorts can be accessed through Hartwig Medical Foundation on approval of a research access request (https://www.hartwigmedicalfoundation.nl/en/data/data-acces-request). As determined in the original publication, NICHE study RNA-seq and DNA-seq data have been deposited into the European Genome–Phenome Archive under accession number EGAS00001004160 and are available on reasonable request for academic use and within the limitations of the provided informed consent. The scRNA-seq data have been deposited at the GEO (GSE216534) and are publicly available. All other data are available from the corresponding author on reasonable request. The GRCh38 primary assembly of the human reference genome was downloaded from Gencode (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_42/GRCh38.primary_assembly.genome.fa.gz) with Gencode's matching v29 annotation file (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_29/gencode.v29.annotation.gtf.gz) for gene expression quantification.

References

1. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558–561 (1993).

2. Germano, G. et al. Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature* **552**, 116–120 (2017).

3. Zaretsky, J. M. et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).

4. Gettinger, S. et al. Impaired HLA class i antigen processing and presentation as a mechanism of acquired resistance to immune checkpoint inhibitors in lung cancer. *Cancer Discov.* **7**, 1420–1435 (2017).

5. Sade-Feldman, M. et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat. Commun.* **8**, 1136 (2017).

6. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).

7. Overman, M. J. et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet. Oncol.* **18**, 1182–1191 (2017).

8. Overman, M. J. et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. *J. Clin. Oncol.* **36**, 773–779 (2018).

9. Chalabi, M. et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat. Med.* **26**, 566–576 (2020).

10. Dolcetti, R. et al. High prevalence of activated intraepithelial cytotoxic T lymphocytes and increased neoplastic cell apoptosis in colorectal carcinomas with microsatellite instability. *Am. J. Pathol.* **154**, 1805–1813 (1999).

11. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).

12. Taube, J. M. et al. Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. *Clin. Cancer Res.* **20**, 5064–5074 (2014).

13. Bicknell, D. C., Kaklamanis, L., Hampson, R., Bodmer, W. F. & Karran, P. Selection for β_2-microglobulin mutation in mismatch repair-defective colorectal carcinomas. *Curr. Biol.* **6**, 1695–1697 (1996).

14. Kloor, M. et al. Immunoselective pressure and human leukocyte antigen class I antigen machinery defects in microsatellite unstable colorectal cancers. *Cancer Res.* **65**, 6418–6424 (2005).

15. Dierssen, J. W. et al. HNPCC versus sporadic microsatellite-unstable colon cancers follow different routes toward loss of HLA class I expression. *BMC Cancer* **7**, 33 (2007).

16. Ijsselsteijn, M. E. et al. Revisiting immune escape in colorectal cancer in the era of immunotherapy. *Br. J. Cancer* **120**, 815–818 (2019).

17. Hughes, E. A., Hammond, C. & Cresswell, P. Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. *Proc. Natl Acad. Sci. USA* **94**, 1896–1901 (1997).

18. Middha, S. et al. Majority of B2M-mutant and -deficient colorectal carcinomas achieve clinical benefit from immune checkpoint inhibitor therapy and are microsatellite instability-high. *JCO Precis. Oncol.* **3**, 1–14 (2019).

19. Groh, V. et al. Human lymphocytes bearing T cell receptor γ/δ are phenotypically diverse and evenly distributed throughout the lymphoid system. *J. Exp. Med.* **169**, 1277–1294 (1989).

20. Silva-Santos, B., Mensurado, S. & Coffelt, S. B. γδ T cells: pleiotropic immune effectors with therapeutic potential in cancer. *Nat. Rev. Cancer* **19**, 392–404 (2019).

21. Halary, F. et al. Control of self-reactive cytotoxic T lymphocytes expressing γδ T cell receptors by natural killer inhibitory receptors. *Eur. J. Immunol.* **27**, 2812–2821 (1997).

22. de Vries, N. L. et al. High-dimensional cytometric analysis of colorectal cancer reveals novel mediators of antitumour immunity. *Gut* **69**, 691–703 (2020).

23. van der Velden, D. L. et al. The Drug Rediscovery protocol facilitates the expanded use of existing anticancer drugs. *Nature* **574**, 127–131 (2019).

24. Danaher, P. et al. Gene expression markers of tumor infiltrating leukocytes. *J. Immunother. Cancer* **5**, 18 (2017).

25. Priestley, P. et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).

26. Duhen, T. et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat. Commun.* **9**, 2724 (2018).

27. Kwon, M. et al. Determinants of response and intrinsic resistance to PD-1 blockade in microsatellite instability-high gastric cancer. *Cancer Discov.* **11**, 2168–2185 (2021).

28. Wu, Y. et al. An innate-like Vδ1$^+$ γδ T cell compartment in the human breast is associated with remission in triple-negative breast cancer. *Sci. Transl. Med.* **11**, eaax9364 (2019).

29. Wu, Y. et al. A local human Vδ1 T cell population is associated with survival in nonsmall-cell lung cancer. *Nat. Cancer* **3**, 696–709 (2022).

30. Maeurer, M. J. et al. Human intestinal Vδ1$^+$ lymphocytes recognize tumor cells of epithelial origin. *J. Exp. Med.* **183**, 1681–1696 (1996).

31. Siegers, G. M., Ribot, E. J., Keating, A. & Foster, P. J. Extensive expansion of primary human gamma delta T cells generates cytotoxic effector memory cells that can be labeled with feraheme for cellular MRI. *Cancer Immunol. Immunother.* **62**, 571–583 (2013).

32. Wu, D. et al. Ex vivo expanded human circulating Vδ1 γδT cells exhibit favorable therapeutic potential for colon cancer. *Oncoimmunology* **4**, e992749 (2015).

33. Almeida, A. R. et al. Delta one T cells for immunotherapy of chronic lymphocytic leukemia: clinical-grade expansion/differentiation and preclinical proof of concept. *Clin. Cancer Res.* **22**, 5795–5804 (2016).

34. Mikulak, J. et al. NKp46-expressing human gut-resident intraepithelial Vδ1 T cell subpopulation exhibits high antitumor activity against colorectal cancer. *JCI Insight* **4**, e125884 (2019).

35. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8$^+$ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).

36. Groh, V., Steinle, A., Bauer, S. & Spies, T. Recognition of stress-induced MHC molecules by intestinal epithelial γδ T cells. *Science* **279**, 1737–1740 (1998).

37. Groh, V. et al. Broad tumor-associated expression and recognition by tumor-derived γδ T cells of MICA and MICB. *Proc. Natl Acad. Sci. USA* **96**, 6879–6884 (1999).

38. Poggi, A. et al. Vδ1 T lymphocytes from B-CLL patients recognize ULBP3 expressed on leukemic B cells and up-regulated by *trans*-retinoic acid. *Cancer Res.* **64**, 9172–9179 (2004).

39. Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22**, 1342–1350 (2016).

40. Cader, F. Z. et al. A peripheral immune signature of responsiveness to PD-1 blockade in patients with classical Hodgkin lymphoma. *Nat. Med.* **26**, 1468–1479 (2020).

41. Germano, G. et al. CD4 T cell-dependent rejection of beta-2 microglobulin null mismatch repair-deficient tumors. *Cancer Discov.* **11**, 1844–1859 (2021).

42. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

43. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689 (2018).

44. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **51**, 411–412 (2019).

45. Huang, M. N. et al. MSIseq: software for assessing microsatellite instability from catalogs of somatic mutations. *Sci. Rep.* **5**, 13321 (2015).

46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

47. Shale, C. et al. Unscrambling cancer genomes via integrated analysis of structural variation and copy number. *Cell Genomics* **2**, 100112 (2022).

48. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

49. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

50. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital

gene expression data. *Bioinformatics* **26**, 139–140 (2010).

51. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

52. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

53. Hall, G. et al. Immunohistochemistry for PMS2 and MSH6 alone can replace a four antibody panel for mismatch repair deficiency screening in colorectal adenocarcinoma. *Pathology* **42**, 409–413 (2010).

54. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

55. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).

56. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).

57. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

58. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at *arXiv* https://arxiv.org/abs/1802.03426 (2018).

59. Ijsselsteijn, M. E., van der Breggen, R., Farina Sarasqueta, A., Koning, F. & de Miranda, N. F. C. C. A 40-marker panel for high dimensional characterization of cancer immune microenvironments by imaging mass cytometry. *Front. Immunol.* **10**, 2534–2534 (2019).

60. Berg, S. et al. ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

61. Ijsselsteijn, M. E., Somarakis, A., Lelieveldt, B. P. F., Höllt, T. & de Miranda, N. Semi-automated background removal limits data loss and normalises imaging mass cytometry data. *Cytometry A* **99**, 1187–1197 (2021).

62. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

63. Somarakis, A., Van Unen, V., Koning, F., Lelieveldt, B. P. F. & Hollt, T. ImaCytE: visual exploration of cellular microenvironments for imaging mass cytometry data. *IEEE Trans. Vis. Comp. Graph.* **27**, 98–110 (2019).

64. van der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

65. Höllt, T. et al. Cytosplore: interactive immune cell phenotyping for large single-cell datasets. *Comput. Graph. Forum* **35**, 171–180 (2016).

66. Dijkstra, K. K. et al. Generation of tumor-reactive T cells by co-culture of peripheral blood lymphocytes and tumor organoids. *Cell* **174**, 1586–1598 (2018).

67. Cattaneo, C. M. et al. Tumor organoid-T-cell coculture systems. *Nat. Protoc.* **15**, 15–39 (2020).

68. Dutta, I., Postovit, L. M. & Siegers, G. M. Apoptosis induced via gamma delta T cell antigen receptor “blocking” antibodies: a cautionary tale. *Front. Immunol.* **8**, 776 (2017).

Acknowledgements

We thank K. C. M. J. Peeters, M. G. Kallenberg-Lantrua, D. Berends-van der Meer and F. A. Holman for their help in collecting and providing samples from patients with colon cancer; the staff at the Flow Cytometry Core Facility of the Leiden University Medical Center for their help with cell sorting; the staff at the Leiden Genome Technology Center for their help

with scRNA-seq; M. Ganesh for help with cell culturing; D. Thommen for discussions; I. S. Rodriguez for the establishment of a *B2M*-knockout organoid line; L. Hoes for initial clinical findings; the staff at the Flow Cytometry Core Facility at the Netherlands Cancer Institute for their support; X. Kong for providing the lentiCRISPR plasmid for *B2M* knockout; the staff at Merus for providing anti-PD-1 antibodies for organoid experiments; and the staff at The Cancer Genome Atlas (TCGA) for providing data used in this manuscript. N.F.C.C.d.M. is funded by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 852832). E.E.V. received funding from the Oncode Institute and Open Targets (Identification of targets modulating lymphocyte-mediated tumour cell killing (project ID: OTAR2061); project leaders, M. Garnett and E.E.V.) and the Josephine Nefkens Foundation. F.K. was supported by the collaboration project TIMID (LSHM18057-SGF) financed by the PPP allowance made available by Top Sector Life Sciences & Health to Samenwerkende Gezondheidsfondsen (SGF) to stimulate public–private partnerships and co-financing by health foundations that are part of the SGF.

Author information

Author notes

1. These authors contributed equally: Natasja L. de Vries, Joris van de Haar, Vivien Veninga, Myriam Chalabi

2. These authors jointly supervised this work: Frits Koning, Noel F. C. C. de Miranda, Emile E. Voest

Authors and Affiliations

1. Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

   Natasja L. de Vries, Marieke E. Ijsselsteijn, Manon van der Ploeg, Jitske van den Bulk, Dina Ruano & Noel F. C. C. de Miranda

2. Department of Immunology, Leiden University Medical Center, Leiden, The Netherlands

Natasja L. de Vries & Frits Koning

3. Department of Molecular Oncology and Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Joris van de Haar, Vivien Veninga, Myriam Chalabi, John B. Haanen, Laurien J. Zeverijn, Birgit S. Geurts, Gijs F. de Wit, Thomas W. Battaglia, Ton N. Schumacher & Emile E. Voest

4. Oncode Institute, Utrecht, The Netherlands

Joris van de Haar, Vivien Veninga, Laurien J. Zeverijn, Birgit S. Geurts, Gijs F. de Wit, Thomas W. Battaglia, Ton N. Schumacher, Lodewyk F. A. Wessels & Emile E. Voest

5. Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Amsterdam, The Netherlands

Joris van de Haar & Lodewyk F. A. Wessels

6. Gastrointestinal Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Myriam Chalabi

7. Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Myriam Chalabi & John B. Haanen

8. Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Jose G. van den Berg

9. Department of Medical Oncology, Leiden University Medical Center, Leiden, The Netherlands

Hans Gelderblom

10. Department of Medical Oncology, Erasmus MC, Rotterdam, The Netherlands

Henk M. W. Verheul

11. Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands

Ton N. Schumacher

12. Faculty of EEMCS, Delft University of Technology, Delft, The Netherlands

Lodewyk F. A. Wessels

Contributions

N.L.d.V., J.v.d.H. and V.V. conceived the study and performed experiments. J.v.d.H. performed genomic and bulk transcriptomic analyses of ICB-naive as well as ICB-treated MMR-d cancers. N.L.d.V. performed scRNA-seq and cell culturing experiments. V.V. performed organoid experiments. N.L.d.V., M.v.d.P. and J.v.d.B. performed cell line reactivity and immune cell killing experiments. N.L.d.V., V.V. and M.v.d.P. performed blocking experiments. M.C. provided tissue sections of patients in the NICHE study, which was designed and coordinated by M.C. under the joint supervision of T.N.S., E.E.V. and J.B.H.; M.E.I. performed IMC experiments. M.E.I., N.F.C.C.d.M., N.L.d.V. and D.R. analysed the IMC data. N.L.d.V. and D.R. analysed the scRNA-seq data. J.G.v.d.B. evaluated histological and immunohistochemical analyses. L.J.Z., B.S.G., G.F.d.W., T.W.B., H.G. and H.M.W.V. designed, coordinated and analysed data from the DRUP study. General scientific coordination was undertaken by J.v.d.H.; L.F.A.W., F.K., N.F.C.C.d.M. and E.E.V. supervised the study; T.N.S. had an advisory role. The manuscript was written by N.L.d.V., J.v.d.H. and V.V. in collaboration

with all of the authors. All of the authors commented on and approved the manuscript.

Corresponding authors

Correspondence to Noel F. C. C. de Miranda or Emile E. Voest.

Ethics declarations

Competing interests

M.C. has performed an advisory role or offered expert testimony for BMS, MSD and NUMAB; has received honoraria from BMS and Roche; and has received financing of scientific research from Roche, BMS and MSD. J.B.H. has received research funding from BMS; and has performed an advisory role for BMS.

Peer review

Peer review information

Nature thanks Luis Diaz Jr and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Association of *B2M* status to clinical, genomic, and transcriptomic characteristics of MMR-d and MMR-p tumours.

a. Biopsy site distribution in the DRUP cohort (n = 71 patients). Colours denote patients' *B2M* status (WT: wildtype,; ALT: altered). Fisher's exact test-based two-sided P-values for enrichment/depletion of *B2M* altered tumours per biopsy site are shown. **b**. Tumour mutational burden *vs B2M* status. Wilcoxon rank sum test-based two-sided P-value is shown. Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and individual data points, respectively. **c**. As **b**, but for clinical benefit to ICB (x-axis). **d**. Volcano plots indicating differential gene expression between B2M-based subgroups (see title) of MMR-d cancers in TCGA COAD (colon adenocarcinoma; n = 57 patients), STAD (stomach adenocarcinoma; n = 60 patients) and UCEC (uterus corpus endometrial carcinoma; n = 122 patients) cohorts. Results were adjusted for tumour type and multiple hypothesis testing (Methods). **e**. Hierarchical clusters of expression of genes significantly (FDR <25%; see Fig. 1D) upregulated in MMR-d $B2M^{MUT}$ *vs* MMR-d $B2M^{WT}$ cancers in the TCGA COAD/STAD/UCEC cohorts. The blue dashed rectangle denotes the Vδ1/3 T cell cluster. **f**. Immune marker gene set expression in $B2M^{WT}$ (pink), and $B2M^{MUT}$ (red) MMR-d cancers in the TCGA COAD/STAD/UCEC cohorts separately or combined (All). Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and individual data points, respectively. Wilcoxon rank sum test-based two-sided P-values are shown. **g**. As **f**, but for MMR-d cancers in the DRUP cohort, for all cancers combined (All), only colorectal cancer (CRC), or all non-CRC cancers (Other). Two-sided P-values were calculated with linear regression adjusting for biopsy site and tumour type (Methods). **h**. Allelic alteration status of *B2M* in the Hartwig cohort of MMR-p cancers. **i**. RNA expression of Vδ1+Vδ3 loci in MMR-p $B2M^{WT}$ (grey), and MMR-p $B2M^{MUT}$ (red) cancers in the Hartwig cohort, stratified per primary tumour location. Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and individual data points, respectively. The linear regression-based, two-sided, primary tumour location-adjusted P-value for association of *B2M* status with Vδ1+Vδ3 loci expression is shown.

Extended Data Fig. 2 Characterization of γδ T cells and other immune cell populations infiltrating MMR-d colon cancers.

a. FACS gating strategy for single, live $CD45^+$ $EpCAM^-$ $CD3^+$ $TCR\gamma\delta^+$ cells from a representative MMR-d colon cancer sample showing sequential

gates with percentages. **b**. Frequencies of positive cells for selected genes across Vδ1 (n = 1927), Vδ2 (n = 860), and Vδ3 (n = 506) cells as percentage of total γδ T cells from each MMR-d colon tumour (n = 5) analysed by single-cell RNA-sequencing. Vδ3 cells were present in two out of five colon cancers. Bars and dots indicate median ± IQR and individual samples, respectively. **c**. Frequencies of marker-positive γδ T cells, $CD56^+$ NK cells, $CD4^+$ T cells, and $CD8^+$ T cells in treatment-naive $\beta2m^-$ (n = 5) MMR-d colon cancers. $CD56^+$ NK cells were present in four out of five $\beta2m^-$ cancer samples. Bars and dots indicate median ± IQR and individual samples, respectively. **d**. Frequencies of $CD103^+CD39^+$ γδ T cells, $CD8^+$ T cells, and $CD4^+$ T cells in treatment-naive $\beta2m^-$ (n = 5) MMR-d colon cancers. Bars indicate median ± IQR and individual samples, respectively.

Extended Data Fig. 3 Distinct clusters of γδ T cells in MMR-d colon cancers by single-cell RNA-sequencing.

a. UMAP embedding showing γδ T cells (n = 4442) isolated from MMR-d colon cancers (n = 5) analysed by single-cell RNA-sequencing. Colours represent the functionally different γδ T cell clusters identified by graph-based clustering and non-linear dimensional reduction. Dots represent single cells. **b**. UMAP embedding of (**a**) coloured by patient ID. Dots represent single cells. **c**. Heatmap showing the normalized single-cell gene expression value (z-score, purple-to-yellow scale) for the top 10 differentially expressed genes in each identified γδ T cell cluster.

Extended Data Fig. 4 Sorting and expansion of γδ T cells from MMR-d colon cancers.

a. FACS gating strategy for single, live $CD3^+$ $TCR\gamma\delta^+$ cells from a representative MMR-d colon cancer sample showing sequential gates with percentages. **b**. Sorting of all γδ T cells from CRC94 (due to the low number of $PD\text{-}1^+$ cells), and of $PD\text{-}1^-$ (blue squares) and $PD\text{-}1^+$ (orange squares) γδ T cells from CRC167, CRC134, CRC96, and CRC154. Dots represent single cells. **c**. Table showing the number of γδ T cells isolated from MMR-d colon cancers (n = 5) at the time of sorting *vs* 3-4 weeks after expansion, and the fold increase thereof. **d**. TCR Vδ chain usage after expansion of γδ T cells

from CRC94 and CRC167 (first row), and at the time of sorting (left panel) as well as after expansion (right panel) of γδ T cells from CRC134, CRC96, and CRC154. PD-1⁻ γδ T cells from CRC154 did not expand in culture. Dots represent single cells.

Extended Data Fig. 5 Phenotype and reactivity of γδ T cells towards cancer cell lines.

a. The percentage of positive cells for the indicated markers on expanded γδ T cells from MMR-d colon cancers (n = 5). **b**. Diagram showing the *B2M* status and surface expression of HLA class I, NKG2D ligands, DNAM-1 ligands, and butyrophilin on cancer cell lines. **c**. Bar plots showing CD137 expression on γδ T cells upon co-culture with cancer cell lines. Medium was used as negative control and PMA/ionomycin as positive control. Bars indicate mean ± SEM. Data from at least two independent experiments except for CRC167, depending on availability of γδ T cells. **d**. Representative flow cytometry plots showing CD137 expression on γδ T cells from a MMR-d colon cancer upon co-culture with cancer cell lines as compared to medium only. Gates indicate percentage of positive γδ T cells. **e**. Bar plots showing the presence of IFNγ in the supernatant upon co-culture of γδ T cells from MMR-d colon cancers (n = 5) with cancer cell lines. Medium as negative control and PMA/ionomycin as positive control are included. Bars indicate mean ± SEM of triplicates. **f**. Bar plots showing the expression of OX40 (first row) and PD-1 (second row) on γδ T cells upon co-culture of γδ T cells with CRC cell lines. PD-1 expression is shown as difference from baseline (medium) condition. Bars indicate mean ± SEM. Data from at least two independent experiments.

Extended Data Fig. 6 Reactivity of γδ T cells towards *B2M*-knockin vs -wildtype cancer cell lines.

a. Flow cytometry gating strategy to validate β2m expression on HCT-15 and LoVo *B2M*-knockin (*B2M*-KI) cell lines. Isotype controls were included as negative control. **b**. Bar plots showing the quantification of killing of HCT-15 *B2M*-KI *vs* wildtype (WT) cells upon co-culture with γδ T cells from MMR-d colon cancers (n = 5) in the presence of a red fluorescent caspase-3/7 reagent. Bars indicate mean ± SEM of two wells with two

images/well. Right panel shows representative time course of apoptosis. **c**. As **b**, but for LoVo *B2M*-KI *vs* WT cells.

Extended Data Fig. 7 Tumour organoid characterization and reactivity assay readout.

a. Flow cytometry gating strategy on PDTO cells for analysis of surface staining. Selected cells were gated on single, live cells before quantification of staining signal. **b**. Histogram representation and count for surface staining of MHC-I, PD-L1, and β2m expression on two PDTO lines $B2M^{WT}$ and $B2M^{KO}$ after IFNγ pre-stimulation. Staining with isotype antibodies for each fluorochrome (PE, APC and FITC) were included as negative control. **c**. Flow cytometry gating strategy on γδ T cell samples for analysis of intracellular staining to test antitumour reactivity upon PDTO stimulation. Lymphocyte population was further gated on single cells, live and $CD3^+$ cells, $γδTCR^+$ cells and $CD8^+$ as well as $CD8^-CD4^-$ cells. Reactivity of the sample was based on $IFNγ^+$ cells of the selected population. **d**. Histogram representation and count for surface staining of NKG2D ligands MICA/B, ULBP1, ULBP2/5/6, ULBP3, and ULBP4 on two PDTO lines $B2M^{WT}$ and $B2M^{KO}$ after IFNγ pre-stimulation. **e**. Flow cytometry gating strategy on γδ T cell samples for analysis of intracellular staining after stimulation with PDTOs in the presence of NKG2D ligand blocking. Lymphocyte population was further gated on single cells, live and $CD3^+$ cells, followed by $γδTCR^+$ and $CD8^+$ as well as $CD8^-$ cells. Reactivity of final population was based on $IFNγ^+$ or $CD107a^+$ cells.

Extended Data Fig. 8 Reactivity of γδ T cells towards cancer cell lines in the presence of NKG2D ligand blocking.

a. Representative flow cytometry plots showing NKG2D expression on γδ T cells from a MMR-d colon cancer upon co-culture with cancer cell lines as compared to medium only. Gates indicate percentage of positive γδ T cells. **b**. Bar plots showing NKG2D expression on γδ T cells from MMR-d colon cancers (n = 5) upon co-culture with cancer cell lines. Medium as negative control and PMA/ionomycin as positive control are included. Bars indicate mean ± SEM. Data from at least two independent experiments except for

CRC167 (SW403, SK-CO-1, K-562), depending on availability of γδ T cells. **c**. Bar plots showing the killing of HT-29 cells upon co-culture with γδ T cells with or without NKG2D ligand blocking. Bars indicate mean ± SEM of two wells with two images/well. **d**. Bar plots showing the expression of CD137 (first row), OX40 (second row), and PD-1 (third row) on γδ T cells upon co-culture of γδ T cells with CRC cell lines with or without NKG2D ligand blocking. PD-1 expression is shown as difference from baseline (medium) condition. Bars and lines indicate mean and similar experiments, respectively. Data from two independent experiments for CD137 and OX40.

Extended Data Fig. 9 Loss of β2m protein expression on tumour cells in *B2M*-mutant MMR-d colon cancers.

Immunohistochemical analysis of β2m protein expression in FFPE tissue from all five $B2M^{MUT}$ MMR-d colon cancers of the NICHE cohort. A $B2M^{WT}$ case (GD02) staining positive for β2m is included as control. Details on the staining procedure can be found in Methods.

Extended Data Fig. 10 Distribution of immune cell populations in *B2M*-mutant and *B2M*-wildtype MMR-d colon cancers pre- and post-ICB treatment in the NICHE trial.

a. Immune marker gene set RNA expression in MMR-d $B2M^{WT}$ (pink), and MMR-d $B2M^{MUT}$ (red) cancers, before (left) and after (right) neoadjuvant ICB. Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and individual data points, respectively. Wilcoxon rank sum test-based two-sided P-values are shown. **b**. Pre- (grey) and post-ICB (orange) RNA expression of KIRs in $B2M^{MUT}$ MMR-d cancers in the NICHE study (n = 5). Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and outliers, respectively. Two-sided P-values were calculated by Wilcoxon rank sum test. **c**. Pre- and post-ICB frequencies of γδ T cells by imaging mass cytometry in $B2M^{WT}$ (n = 5) and $B2M^{MUT}$ (n = 5) MMR-d colon cancers (corresponding to Fig. 4b). Lines indicate paired samples, dots represent individual samples. Wilcoxon rank sum test-based two-sided P-values are shown. **d**. Imaging mass cytometry-based frequencies of intraepithelial γδ T cells, $CD56^{+}$ NK cells, $CD4^{+}$ T cells, and $CD8^{+}$ T cells in ICB-naive $B2M^{WT}$

(n = 5) and *B2M*^{MUT} (n = 5) MMR-d colon cancers. Bars and dots indicate median ± IQR and individual samples, respectively. Wilcoxon rank sum test-based two-sided P-values are shown. **e**. Cell counts (cells/mm^2) of immune populations from the imaging mass cytometry of *B2M*^{WT} (n = 5) and *B2M*^{MUT} (n = 5) MMR-d colon cancers upon ICB treatment. Colour bar is scaled per immune population. **f**. Representative images of the detection of cytotoxic (granzyme B$^+$), tissue-resident (CD103$^+$), activated (CD39$^+$), proliferating (Ki-67$^+$), and PD-1$^+$ γδ T cells by imaging mass cytometry in a *B2M*^{MUT} MMR-d colon cancer upon ICB treatment. **g**. Frequencies of marker-positive γδ T cells, CD56$^+$ NK cells, CD4$^+$ T cells, and CD8$^+$ T cells in the sole *B2M*^{MUT} MMR-d colon cancer that contained cancer cells upon ICB treatment. **h**. Frequencies of CD103$^+$CD39$^+$ γδ T cells, CD8$^+$ T cells, and CD4$^+$ T cells in the sole *B2M*^{MUT} MMR-d colon cancer that contained cancer cells upon ICB treatment.

Supplementary information

Supplementary Information

A guide to the Supplementary Information, and Supplementary Tables 4–7.

Reporting Summary

Supplementary Table 1

Characteristics of clinical samples from 71 patients with MMR-d cancers from the DRUP study.

Supplementary Table 2

Immune marker gene sets for bulk RNA-seq analyses.

Supplementary Table 3

Characteristics of clinical samples from 2,256 patients with MMR-p cancers from the Hartwig database.

Supplementary Video 1

Killing of CRC cells after co-culture with γδ T cells. Killing of HCT-15 cells (red) by γδ T cells (V$\delta1^+$) from a MMR-d colon cancer in the presence of caspase-3/7 (green). γδ T cells were added at $T = 6$ h, and images were taken until 12 h after ($T = 18$ h). Cancer cell apoptosis is visualized in yellow.

Supplementary Video 2

Pronounced cancer cell apoptosis after co-culture with PD-1^+ (V$\delta1^+$) as compared to PD-1^- γδ T cells. Killing of HCT-15 cells (red) by PD-1^+ (V$\delta1^+$, left) as compared to PD-1^- (V$\delta2^+$, right) γδ T cells from a MMR-d colon cancer in the presence of caspase-3/7 (green). γδ T cells were added at $T = 5$ h, and images were taken until 12 h after ($T = 17$ h). Cancer cell apoptosis is visualized in yellow.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

de Vries, N.L., van de Haar, J., Veninga, V. *et al.* γδ T cells are effectors of immunotherapy in cancers with HLA class I defects. *Nature* **613**, 743–750 (2023). https://doi.org/10.1038/s41586-022-05593-1

- Received: 16 July 2021
- Accepted: 24 November 2022
- Published: 11 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05593-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

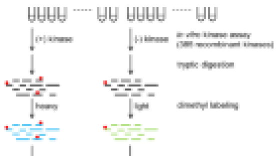
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

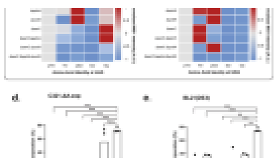
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05593-1



Large-scale Discovery of Substrates of the Human Kinome

19 July 2019

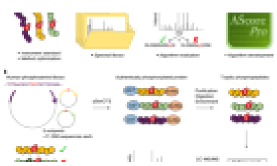
Naoyuki Sugiyama, Haruna Imamura & Yasushi Ishihama



Enhanced access to the human phosphoproteome with genetically encoded phosphothreonine

24 November 2022

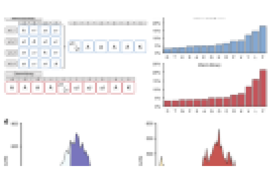
Jack M. Moen, Kyle Mohler, … Jesse Rinehart



A multi-purpose, regenerable, proteome-scale, human phosphoserine resource for phosphoproteomics

24 October 2022

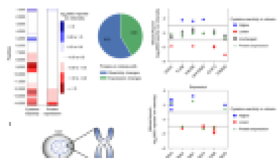
Brandon M. Gassaway, Jiaming Li, … Steven P. Gygi



[Dissecting the sequence determinants for dephosphorylation by the catalytic subunits of phosphatases PP1 and PP2A]

17 July 2020

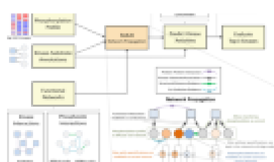
Bernhard Hoermann, Thomas Kokot, … Maja Köhn



[Global profiling of phosphorylation-dependent changes in cysteine reactivity]

28 February 2022

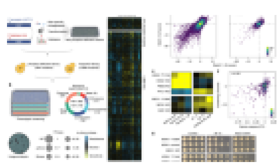
Esther K. Kemper, Yuanjin Zhang, … Benjamin F. Cravatt



[Robust inference of kinase activity using functional networks]

19 February 2021

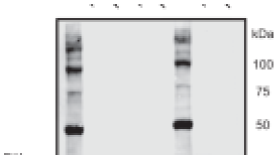
Serhan Yılmaz, Marzieh Ayati, … Mehmet Koyutürk



[High-throughput functional characterization of protein phosphorylation sites in yeast]

18 October 2021

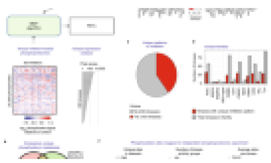
Cristina Viéitez, Bede P. Busby, … Pedro Beltrao



[Phosphorylation Changes in Response to Kinase Inhibitor H89 in PKA-Null Cells]

26 February 2019

Kavee Limbutara, Andrew Kelleher, … Mark A. Knepper



[Reconstructing kinase network topologies from phosphoproteomics data reveals cancer-associated rewiring]

20 January 2020

Maruan Hijazi, Ryan Smith, … Pedro R. Cutillas

- Article
- Open Access
- Published: 11 January 2023

An atlas of substrate specificities for the human serine/threonine kinome

- Jared L. Johnson,

- Tomer M. Yaron,
- Emily M. Huntsman,
- Alexander Kerelsky,
- Junho Song,
- Amit Regev,
- Ting-Yu Lin,
- Katarina Liberatore,
- Daniel M. Cizin,
- Benjamin M. Cohen,
- Neil Vasan,
- Yilun Ma,
- Konstantin Krismer,
- Jaylissa Torres Robles,
- Bert van de Kooij,
- Anne E. van Vlimmeren,
- Nicole Andrée-Busch,
- Norbert F. Käufer,
- Maxim V. Dorovkov,
- Alexey G. Ryazanov,
- Yuichiro Takagi,
- Edward R. Kastenhuber,
- Marcus D. Goncalves,
- Benjamin D. Hopkins,
- Olivier Elemento,
- Dylan J. Taatjes,
- Alexandre Maucuer,
- Akio Yamashita,
- Alexei Degterev,
- Mohamed Uduman,
- Jingyi Lu,
- Sean D. Landry,
- Bin Zhang,
- Ian Cossentino,
- Rune Linding,
- John Blenis,
- Peter V. Hornbeck,
- Benjamin E. Turk ,

- Michael B. Yaffe &
- …
- Lewis C. Cantley

Show authors

Nature **volume 613**, pages 759–766 (2023)

- 33k Accesses
- 374 Altmetric
- Metrics details

Abstract

Protein phosphorylation is one of the most widespread post-translational modifications in biology[1,2]. With advances in mass-spectrometry-based phosphoproteomics, 90,000 sites of serine and threonine phosphorylation have so far been identified, and several thousand have been associated with human diseases and biological processes[3,4]. For the vast majority of phosphorylation events, it is not yet known which of the more than 300 protein serine/threonine (Ser/Thr) kinases encoded in the human genome are responsible[3]. Here we used synthetic peptide libraries to profile the substrate sequence specificity of 303 Ser/Thr kinases, comprising more than 84% of those predicted to be active in humans. Viewed in its entirety, the substrate specificity of the kinome was substantially more diverse than expected and was driven extensively by negative selectivity. We used our kinome-wide dataset to computationally annotate and identify the kinases capable of phosphorylating every reported phosphorylation site in the human Ser/Thr phosphoproteome. For the small minority of phosphosites for which the putative protein kinases involved have been previously reported, our predictions were in excellent agreement. When this approach was applied to examine the signalling response of tissues and cell lines to hormones, growth factors, targeted inhibitors and environmental or genetic perturbations, it revealed unexpected insights into pathway complexity and compensation. Overall, these studies reveal the intrinsic substrate specificity

of the human Ser/Thr kinome, illuminate cellular signalling responses and provide a resource to link phosphorylation events to biological pathways.

Main

Phosphorylation of proteins at serine, threonine, tyrosine and histidine residues controls nearly every aspect of eukaryotic cellular function[1,2,5,6]. Misregulation of protein kinase signalling commonly results in human disease[7]. Deciphering the cellular roles of any protein kinase requires the elucidation of its downstream effector substrates. However, the majority of kinase–substrate relationships that have been published to date involve a relatively small number of well-studied protein kinases, while few, if any, substrates have been identified for the majority of the approximately 300 human protein Ser/Thr kinases within the human kinome[8,9,10]. This lack of knowledge of kinase–substrate relationships limits the interpretation of large mass spectrometry (MS)-based phosphoproteomic datasets, which to date have collectively reported over 200,000 Ser and Thr phosphorylation sites on human proteins[3,4,11,12,13]. The specific kinases that are responsible for these phosphorylation events have been reported for less than 4% of these sites[3], substantially limiting the understanding of cellular phosphorylation networks.

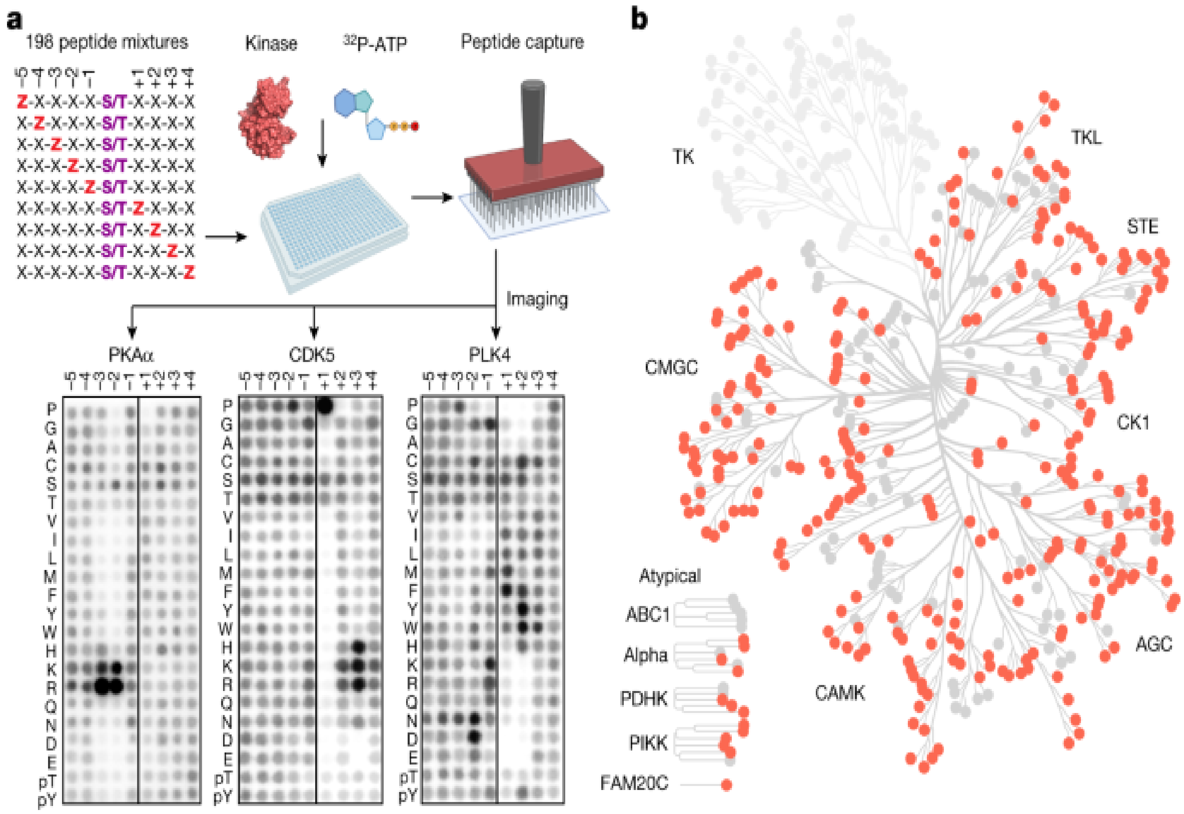
Well-studied Ser/Thr kinases are generally known to recognize specific amino acid residues at multiple positions surrounding the site of phosphorylation[14,15,16,17]. This short linear motif, which is characteristic of a given protein kinase, ensures fidelity in signalling pathways regulating phosphorylation at a given Ser or Thr residue. Knowledge of kinase-recognition motifs can facilitate the discovery of new substrates, for example, by scanning phosphoproteomics data for matching sequences. However, to date, phosphorylation-site sequence motifs are known for only a subset of the human protein Ser/Thr kinome. We have previously described combinatorial peptide library screening methods that enable the rapid determination of specificity for individual kinases based on phosphorylation of peptide substrates[18,19]. Here we apply those methods to experimentally determine the optimal substrate specificity for the large majority of the human Ser/Thr kinome, characterize the relationship between

kinases on the basis of their motifs, and computationally use these data to identify the protein kinases that are likely to phosphorylate any site identified using MS or other techniques. Finally, we show how this information can be applied to capture complex changes in signalling pathways in cells and tissues after genetic, pharmacological, metabolic and environmental perturbations.

Profiling kinase substrate specificity

Substrate-recognition motifs across the human Ser/Thr kinome were determined by performing a positional scanning peptide array (PSPA) analysis. We used a previously reported combinatorial peptide library that systematically substitutes each of 22 amino acids (20 natural amino acids plus phosphorylated Thr and phosphorylated Tyr) at nine positions surrounding a central phospho-acceptor position containing equivalent amounts of Ser and Thr[19] (Fig. 1a). Using purified recombinant kinase preparations, we successfully obtained phosphorylation-site motifs for 303 Ser/Thr kinases, covering every branch of the human Ser/Thr kinase family tree as well as a collection of atypical protein kinases (Fig. 1b, Supplementary Fig. 1 and Supplementary Tables 1 and 2). Profiling of the large majority of these kinases, including 83 understudied 'dark' kinases, was lacking[8].

Fig. 1: Profiling the substrate specificity of the human serine/threonine kinome.

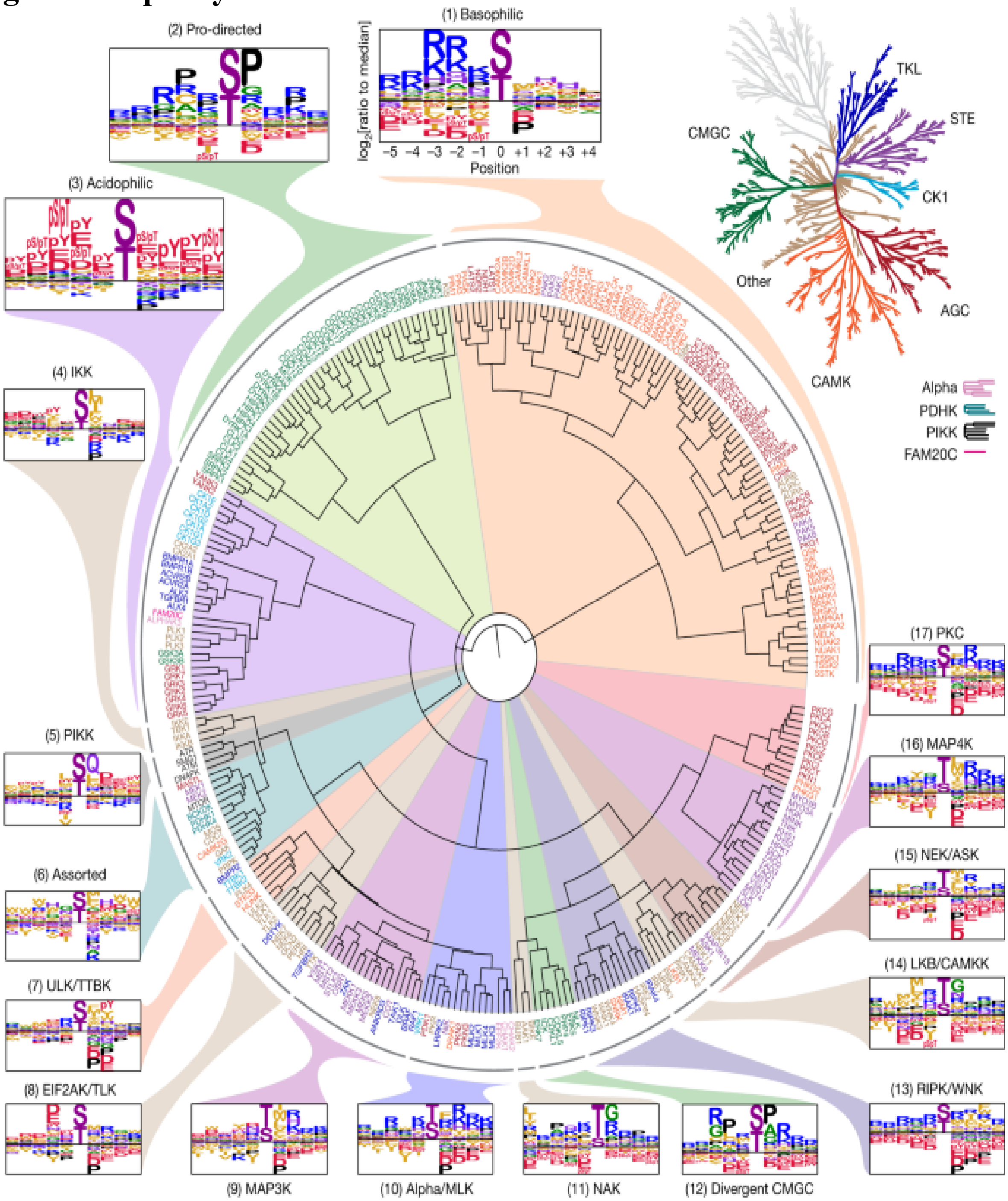


a, Experimental workflow for the PSPA analysis and representative results. The schematic was created using BioRender. Z denotes fixed positions containing one of the 20 natural amino acids, or either phosphorylated Thr (pThr) or phosphorylated Tyr (pTyr). X denotes unfixed positions containing randomized mixtures of all natural amino acids except Ser, Thr and Cys. Darker spots indicate preferred residues. **b**, Dendrogram of the human protein kinome, highlighting the Ser/Thr kinases analysed in this study.

Position-specific scoring matrices (PSSMs) derived from quantified PSPA data were analysed by hierarchical clustering to compare kinase substrate motifs across the kinome (Fig. 2 and Supplementary Table 2). As expected, kinases sharing substantial sequence identity displayed a high degree of similarity in their optimal substrate motifs. However, we found many cases in which clustering by PSSM did not strictly recapitulate the evolutionary phylogenetic relationships between kinases inferred from their primary sequences (Fig. 2). Instead, members of most of the major kinase groups were distributed throughout the dendrogram, reflecting numerous examples in which kinases with low overall sequence identity have converged to

phosphorylate similar optimal sequence motifs. For example, we found that a number of distantly related kinases (in the YANK, CK1 and 2, GRK and TGFβ receptor families) converged to phosphorylate similar sequence motifs despite their disparate locations on the kinome tree (Fig. 2 (cluster 3)).

Fig. 2: Phosphorylation-site motif tree of the human Ser/Thr kinome.



Hierarchical clustering of 303 Ser/Thr kinases on the basis of their amino acid motif selectivity (PSSMs). Kinase names are colour labelled according to their phylogenetic relationships (top right)[2].

Overall inspection of sequence motifs associated with various branches of the motif-based dendrogram revealed that approximately 60% of the Ser/Thr kinome could be represented by simple assignment to one of three previously observed motif classes: selectivity for basic residues N terminal to the phosphorylation site (cluster 1; Fig. 2); directed by a proline residue at the +1 position (cluster 2); or a general preference for negatively charged (acidic and phosphorylated) residues at multiple positions (cluster 3)[15,20,21]. Notably, more than half of all of the reported phosphorylation sites observed by MS could be assigned to one of these three signatures (Extended Data Fig. 1). However, each of these motif classes could be further subcategorized on the basis of selectivity both for and against distinct sets of residues at other positions, reflecting considerable diversity within these clusters (Extended Data Figs. 2–4). The remaining approximately 40% of the Ser/Thr kinome comprised many smaller groups that displayed unique sequence determinants (Fig. 2; clusters 4–17). For example, motifs for the PIKK family kinases (ATM, ATR, DNA-PK and SMG1) clustered into a group that primarily selected a Gln residue at the +1 position (cluster 5), consistent with previous studies[22,23]. Notably, several clusters displayed shared consensus motifs that have not been well recognized previously, such as the group including the IRAK, IRE, WNK, SNRK and RIP kinases (cluster 13), of which the substrate motifs contained basic residues both N and C terminal to the phosphorylation site with dominant selection for aromatic residues at the +3 position. As another example, the kinases LKB1, CAMKK, PINK1 and PBK (cluster 14) primarily recognized hydrophobic residues N terminal to the phosphorylation site in combination with selection for turn-promoting residues (Gly or Asn) at the +1 position. Structural modelling of kinase–peptide complexes revealed complementary features within the kinase catalytic clefts that are probably responsible for the recognition of these motifs (Extended Data Fig. 5a,b).

An important and less generally recognized feature that dominated the clustering was strong negative selection against either positively or negatively charged residues at distinct positions within a motif, suggesting

that electrostatic filtering strongly influences kinase substrate selection throughout the kinome[24]. We identified additional classes of amino acids, such as hydrophobic residues, that are selected against by a variety of kinases. These trends suggest that substrate avoidance has a fundamental role in dictating correct kinase–substrate interactions[25,26].

Unexpectedly, we observed that many kinases (129 out of 303) selected either a phosphorylated Thr or a phosphorylated Tyr as the preferred amino acid in at least one position within the motif (Supplementary Fig. 1; where selectivity for phosphorylated Ser was assumed to be equivalent to phosphorylated Thr). In addition to kinases of which the dependence on phospho-priming was previously known (GSK3, CK1 and CK2 families; cluster 3), this phenomenon was particularly evident for the GRK- and YANK-family kinases (Extended Data Fig. 4), both of which have complementary basic residues within their catalytic domains (Extended Data Fig. 5c,d). Notably, individual GRK-family members showed unique and specific selection for the location of the phosphorylated Thr or phosphorylated Tyr residue within their substrate peptides. GRKs are best known for their role in the desensitization of G-protein-coupled receptors (GPCRs), whereby multisite phosphorylation induces the binding of arrestin proteins to inhibit signalling[27,28]. Our findings suggest that the capacity for only seven GRKs to differentially regulate 800 distinct GPCRs probably involves a complex interplay between initial sequence-specific phospho-priming of GPCRs by other Ser/Thr and Tyr kinases, followed by a second level of specificity resulting from GRK-dependent phosphorylation and subsequent recognition by a small number of β-arrestins.

Features of substrate-recognition motifs across the entire kinome could be structurally rationalized on the basis of the presence of specificity-determining residues at particular positions within the kinase catalytic domain[29,30,31,32], leading to both expected and unexpected findings. For example, we found that half of the kinases display some degree of selectivity for either a Ser or a Thr as the phospho-acceptor residue (Extended Data Figs. 6 and 7). Consistent with our previously published observations[33], Ser or Thr phospho-acceptor site preference strongly correlated with the identity of the ‘DFG+1’ residue (that is, the residue immediately after the canonical Asp-Phe-Gly motif within the kinase activation loop), with bulky residues

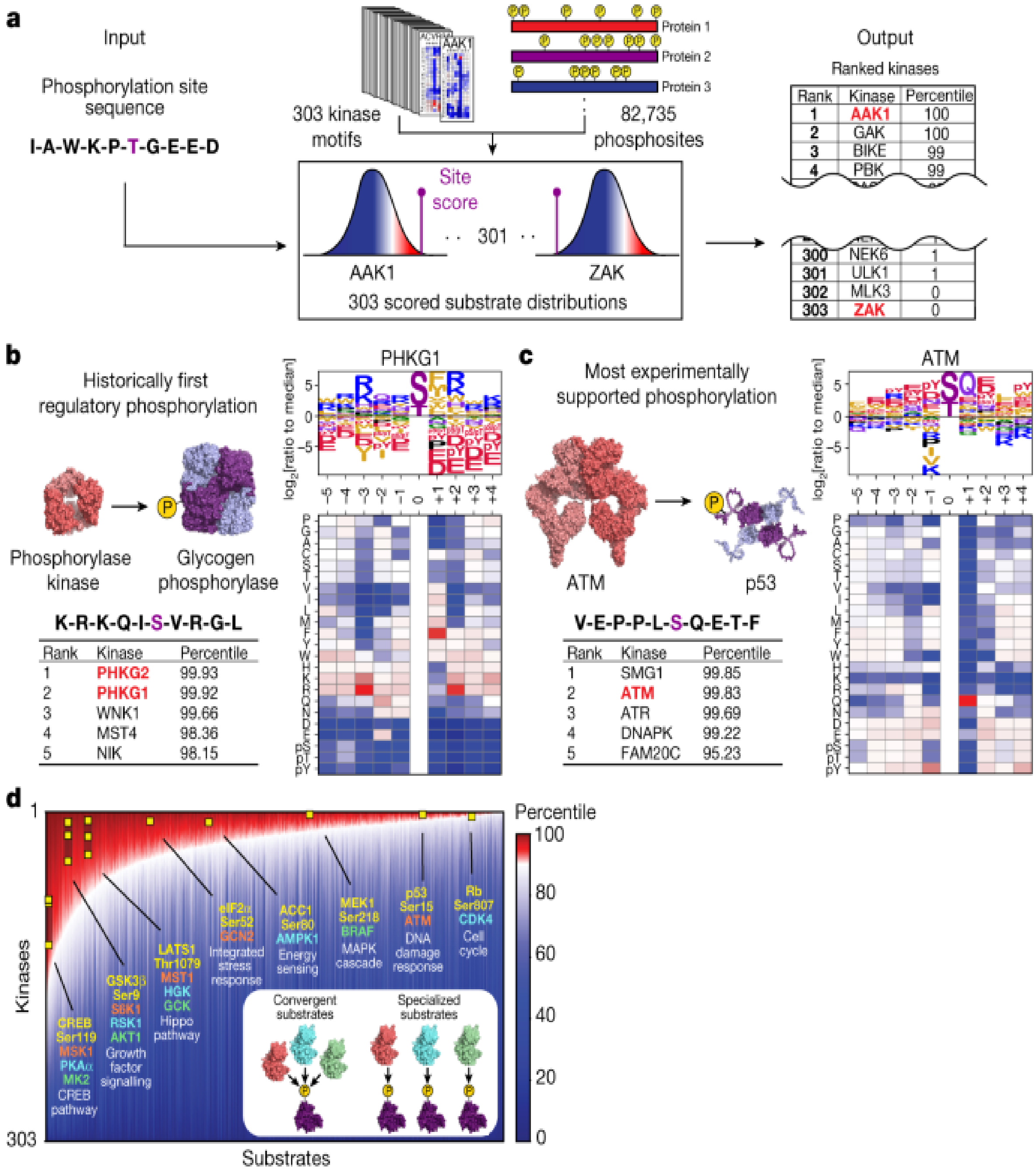
(Phe, Trp, Tyr) at this position in Ser-selective protein kinases and β-branched residues (Val, Ile, Thr) at this position in Thr-selective kinases. However, for some DFG+1 residues, Ser versus Thr selectivity was unexpectedly context dependent. For example, a Leu residue at the DFG+1 position was observed in both Ser-selective and dual-specificity kinases, whereas a DFG+1 Ala residue resulted in a preference for Thr phosphorylation in the context of some kinases (for example, the mitogen-activated protein kinase kinase kinases (MAP3Ks)), but a preference for Ser specificity in others (the IκB kinases). These observations, notable only within the context of the complete Ser/Thr kinome, indicate that additional residues beyond the previously established DFG+1 position can influence Ser/Thr specificity in a context-dependent manner.

Annotation of the human phosphoproteome

Comprehensive knowledge of human Ser/Thr kinase specificity has the potential to 'deorphanize' the large number of reported phosphorylation sites with no associated kinase. To do so, we generated a kinome-wide annotation of the human Ser/Thr phosphoproteome, comprising a previously curated set of 82,735 sites that were detected using high-throughput MS[4] plus an additional 7,017 sites that were identified using only low-throughput methods[3]. Using approaches adapted from previously published research, we computationally ranked these 89,752 sites against each Ser/Thr kinase motif[34,35] (Fig. 3a and Supplementary Table 3). Notably, approximately 99% of these phosphorylation sites ranked favourably for at least one kinase that we profiled (that is, the site scored in the top 10% of all sites in the human phosphoproteome for that kinase). These annotations were strongly concordant with sites for which protein kinases involved have been previously identified. For phosphorylation sites of which the upstream kinase has been previously verified by at least three independent reports, encompassing 969 sites and over one third of the kinome, our motif-based approach yielded a median percentile of 95% (that is, the reported site received a higher score than 95% of all putative phosphorylation sites in the phosphoproteome for its established kinase) (Extended Data Fig. 8a). Furthermore, when we back-mapped the motifs of all 303 profiled kinases onto the literature-reported phosphorylation sites, our approach yielded a median reported kinase percentile of 92% (that is, the reported kinase scored

more favourably than 92% of all profiled kinases in our atlas for its established substrate) (Extended Data Fig. 9a). These rankings further improved when we considered kinase–substrate pairs with higher numbers of previous reports (Extended Data Figs. 8 and 9), suggesting that, in a large majority of cases, the linear sequence context of phosphorylation sites contributes substantially to kinase–substrate relationships.

Fig. 3: Phosphorylation motifs for the human Ser/Thr kinome enable comprehensive scoring and annotation of the human phosphoproteome.



a, Schematic of the substrate-scoring process[4]. **b**, Results for Ser15 on glycogen phosphorylase alongside PSSM and the substrate motif logo of its established kinase glycogen phosphorylase kinase. **c**, The results for Ser15 of p53 alongside its established kinase ATM. **d**, Annotation of the human Ser and Thr phosphoproteome by percentile scores from 303 Ser/Thr kinases performed as shown in **a**. A total of 89,752 phosphorylation sites that were identified using high-throughput approaches[4] and/or low-throughput

approaches[3] were sorted along the *x*-axis by their numbers of kinases with percentile scores higher than 90. On the *y*-axis, kinase percentile scores were sorted by rank separately for each site and represented in the heat map. Examples of well-studied kinase–substrate relationships are highlighted (yellow squares). Inset: phosphorylation sites on the left end of the plot scored favourably for many kinases, whereas sites on the right end scored favourably for fewer kinases.

Notably, motif predictions alone successfully identified numerous prominently studied kinase–substrate relationships. For example, phosphorylase kinases PHKG1 and PHKG2 emerged as the top two hits (out of 303 kinases) for phosphorylating Ser15 of glycogen phosphorylase (Fig. 3b). This phosphoregulatory event, the very first to be discovered[36], opened up the entire field of phosphorylation-dependent signal transduction. The most highly cited kinase–substrate interaction reported to date is the phosphorylation of the tumour suppressor p53 at Ser15 by the DNA-damage-activated kinase ATM, which scored among the top-ranking kinases associated with that site (Fig. 3c). Notably, other kinases reported to phosphorylate the same site—ATR, SMG1 and DNAPK—scored within the top four predicted kinases[3].

Our approach could also correctly identify kinases for phosphorylation events driven by substrate co-localization or non-catalytic docking interactions, for which we expected less dependence on the phosphorylation-site motifs of their kinases. For example, we correctly identified both the mitochondrial-localized phosphorylation of pyruvate dehydrogenase by the pyruvate dehydrogenase kinases (Extended Data Fig. 10a) and the docking-driven phosphorylation of the MAP kinase ERK by MEK[37] (Extended Data Fig. 10b). Notably, the phosphorylation site on ERK was selected against by nearly every human protein kinase that we profiled except for MEK, explaining how ERK can be exclusively regulated by MEK while avoiding phosphorylation by the kinome at large. Finally, our approach could tease apart kinase subfamilies with similar motifs and correctly assign them to their established substrates. For example, we could distinguish between the CDK family kinases that assume classical roles in cell cycle progression (that is, CDK1, CDK2, CDK3, CDK4 and CDK6) and the subset of CDKs

that govern gene transcription (that is, CDK7, CDK8, CDK9, CDK12, CDK13 and CDK19)[38,39] (Extended Data Fig. 11).

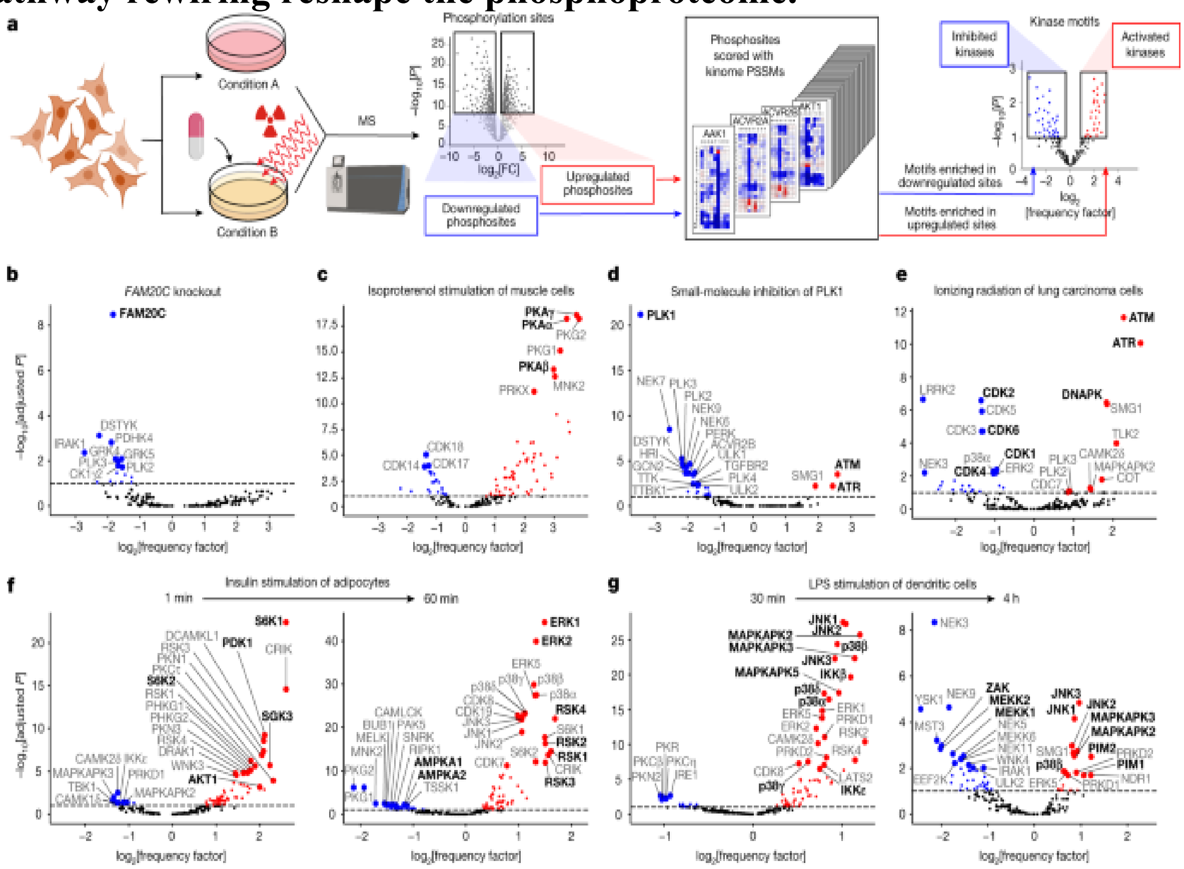
Functional annotation of the human phosphoproteome enabled us to examine global trends in kinase–substrate interactions. We found that most phosphorylation sites could be assigned to a small number of putative kinases (that is, BRAF–MEK1, ATM–p53 and CDK4–Rb; Fig. 3d and Supplementary Table 3). However, approximately one-third of all sites lacked unique negative sequence-discriminating features and, instead, matched well to the optimal phosphorylation motifs for a greater number of kinases[21,40,41,42] (that is, Ser119 of CREB, Ser9 of GSK3B and Thr1079 of LATS1; Fig. 3d). This could suggest the importance of other kinase-determining factors (scaffolds, localization and so on) for proper kinase substrate recognition, or may indicate that these specific phosphorylation sites are points of convergence for multiple signalling pathways. For example, cAMP response element binding protein (CREB) is canonically phosphorylated at Ser119 by cAMP-dependent protein kinase (PKA); however, numerous previous reports demonstrate that a broad range of cellular stimuli and drug perturbations impinge on the phosphorylation of this site by no less than ten distinct kinases[3]. Taken together, these findings suggest that the presence of negative-selectivity elements flanking a putative phosphorylation site can be used to insulate a substrate from inappropriate phosphorylation by dozens of related kinases, whereas the absence of such negative selectivity can enable protein kinases in distinct pathways to converge on the same target.

Motif-enrichment analysis

Cell signalling networks are complex and dynamic. Perturbation of kinase signalling pathways by genetic manipulations, treatment with growth factors and ligands, environmental stress or small-molecule inhibitors reshapes the phosphoproteome through both direct and indirect effects as a consequence of secondary signalling responses and/or off-target effects from the experimental treatment[43]. Owing to the interconnected and dynamic nature of phosphorylation networks, distinguishing between initial signalling events and those that result from the subsequent activation of additional signalling pathways is a common and challenging problem. We reasoned that kinases

underlying both primary and secondary phosphorylation events could potentially be revealed by a global motif-based analysis of changes in the corresponding phosphoproteome. To test this idea, we used publicly available MS datasets from cells collected in the absence or presence of various perturbations and scored all phosphorylation sites with our atlas of Ser/Thr kinase motifs. Kinase motifs that were significantly enriched or depleted after experimental treatment were then represented as volcano plots of motif frequencies and adjusted *P* values (Fig. 4a).

Fig. 4: Global motif analysis reveals how kinase perturbations and pathway rewiring reshape the phosphoproteome.



a, Workflow of the motif enrichment analysis of phosphoproteomics data. The schematic was created using BioRender. **b**–**g**, Results from published datasets. **b**, Conditioned medium of HepG2 cells after genetic deletion of *FAM20C*[44]. **c**, Cultured myotubes after 30 min treatment with 2 µM isoproterenol[45]. **d**, HeLa cells after mitotic arrest by treatment for 45 min with 0.1 µM PLK1 inhibitor BI 2536 (ref. [46]). **e**, A549 cells 2 h after exposure to 6 Gy of ionizing radiation[49]. **f**, 3T3-L1 adipocytes after serum

starvation and then 1 min and 60 min treatment with 100 nM insulin[54]. **g**, C57BL/6J mouse bone-marrow-derived dendritic cells after 30 min and 4 h treatment with 100 ng ml^{-1} lipopolysaccharide (LPS)[55]. The enrichments in **b**–**g** were determined using one-sided exact Fisher's tests and corrected for multiple hypotheses using the Benjamini–Hochberg method. Fully annotated versions of these plots are presented in Supplementary Fig. 2.

Using this approach, we found that sequence motifs corresponding to the most direct target of a genetic or chemical perturbation were among the most significantly regulated, as seen, for example, for the genetic deletion of the secreted primordial casein kinase FAM20C (Fig. 4b). When quantitative phosphoproteomics data from HepG2 cells lacking FAM20C[44] were analysed using our kinome-wide dataset, the most downregulated kinase-recognition motif corresponded to that of FAM20C. Similarly, when skeletal-muscle-like myotube cells were stimulated for 30 min with isoproterenol[45], the most upregulated phosphorylation motifs corresponded to multiple isoforms of cAMP-dependent protein kinase (PKA)—canonical effector kinases downstream of the β_1 and β_2 adrenergic receptors (Fig. 4c). Notably, PKA motifs are highly similar to those of several other basophilic kinases, yet we could identify their enrichment in this scenario. Moreover, our comprehensive Ser/Thr kinome motif collection elucidated secondary signalling events in a dataset from HeLa cells arrested in mitosis using the PLK1 inhibitor BI 2536 (Fig. 4d)[46]; here, in addition to observing a notable downregulation of substrates containing the optimal PLK1 motif, we also noted upregulation of substrates phosphorylated by ATM and ATR. This finding is in good agreement with previous reports that PLK1 can suppress DNA damage signalling in mitotic cells[47,48].

Our motif-based analysis could also be used to reveal key signalling events resulting from more complex interventions. For example, we examined phosphoproteomic data from A549 cells treated with 6 Gy of ionizing radiation[49] (Fig. 4e). Our analysis revealed the up- and downregulation of numerous signalling pathways, including upregulation of canonical kinases involved in the DNA-damage response (ATM, ATR, DNA-PK) and downregulation of canonical kinases involved in cell cycle progression (CDK1, CDK2, CDK4 and CDK6), consistent with G1/S and G2/M arrest.

Furthermore, we found up- and downregulation of less-appreciated DNA-damage-responsive kinases (MAPKAPK2[50,51], PLK3[52] and LRRK2[53]).

The full collection of Ser/Thr kinome motifs also enabled the temporal dynamics of signalling to be resolved from time-resolved phosphoproteomic datasets. For example, motif-based analysis of phosphoproteomic data from insulin-treated 3T3-L1 adipocytes[54] revealed rapid activation of the phosphoinositide 3-kinase signalling pathway within 1 min after insulin stimulation followed by subsequent activation of the MAPK pathway, together with downregulation of AMP-activated protein kinases within 60 min (Fig. 4f). Similarly, phosphoproteomic data analysis from lipopolysaccharide-stimulated dendritic cells[55] suggested marked upregulation at 30 min of a set of stress-activated kinases including the IKKs, JNK and p38 MAPKs, along with the MAPKAPK family of p38 effector kinases. This was followed within 4 h by the subsequent upregulation of the PIM kinases and suppression of the MAPKs in parallel with the downregulation of their upstream MAPK3Ks (MEKK1, MEKK2 and ZAK)[56], suggestive of a negative-feedback loop (Fig. 4g). Thus, comprehensive motif-based approaches, when applied to time-resolved phosphoproteomics experiments, can decipher the distinct temporal dynamics of different groups of kinases.

Discussion

Here we present the full spectrum of substrate motifs of the human serine/threonine kinome and provide an unbiased comprehensive framework to further explore their cellular functions. Globally, these motifs are substantially more diverse than expected, suggesting a broader substrate repertoire of the kinome. Hierarchical clustering of this dataset reorganized the kinome into at least 38 motif classes and introduced several shared motif features (Fig. 2 and Extended Data Figs. 2–4).

The Ser/Thr kinases that we profiled were, almost without exception, strongly discriminatory against specific motif features. These findings suggest that fidelity in kinase signalling pathways is largely achieved through selective pressure on substrates to avoid phosphorylation by the majority of irrelevant kinases, and that this may occur by tuning the amino

acid sequences surrounding the phosphorylation sites to be disfavoured by non-cognate kinases. As this negative selection contributes substantially to proper substrate recognition, accurate identification of kinase–substrate relationships requires a comprehensive knowledge of kinase phosphorylation motifs—not only for an individual kinase of interest, but also for all other kinases in the human kinome that might compete for the same substrate pool.

When this kinome-wide dataset was used to predict the specific kinases that are responsible for substrate phosphorylation solely based on the amino acid sequence surrounding the phosphorylation site, the results were highly accurate at identifying correct kinase–substrate relationships, even without knowledge of tissue specificity, scaffolding effects or subcellular localization. Including such additional information will probably further improve these predictive approaches[57,58]. A limitation of using first-order peptide arrays in these experiments is that they do not directly measure the contributions of interpositional contacts within the substrate peptides, which we have previously shown can affect substrate selection for some tyrosine kinases[32], albeit less so for Ser/Thr kinases[59]. Moreover, we were unable to differentiate between positional selection of Ser or Thr residues and direct phosphorylation of neighbouring residues (for example, peptides containing more than one phospho-acceptor). Structural modelling approaches guided by kinase substrate motif data will potentially decipher this additional information to further improve predictions[60,61].

The examination of MS phosphoproteomic datasets using this global collection of motifs yielded potential biological insights and putative kinase substrates (Fig. 4). For example, in cells undergoing exposure to ionizing radiation (Fig. 4e), ATM was predicted to target 37 of the phosphorylation sites that were upregulated, most of which have never been associated as substrates for ATM (Supplementary Table 4). As the application of phosphoproteomics to human clinical samples and disease model systems continues to advance, our comprehensive motif-based approach will be uniquely equipped to unravel the complex signalling that underlies human disease progressions, mechanisms of cancer drug resistance, dietary interventions and other important physiological processes. In summary, we

foresee that this will provide a valuable resource for a broad spectrum of researchers who study signalling pathways in human biology and disease.

Methods

Cell lines

Expi293 (Thermo Fisher Scientific) and HEK293T (ATCC) cells were obtained directly from vendors that perform short tandem repeat genotyping for authentication of human cells and were verified to be mycoplasma-free. Sf9 insect cells were obtained from Thermo Fisher Scientific.

Plasmids

For expression and purification from bacteria, the DNA sequences for the human Ser/Thr kinases, binding partners and chaperones listed below were codon-optimized for expression in *Escherichia coli* using GeneSmart prediction software (GenScript). Optimized coding sequences were synthesized as gBlocks (Integrated DNA Technologies) carrying 16 bp overhangs at the 5′ and 3′ ends to facilitate in-fusion cloning (Clontech) into pET expression vectors (EMD Millipore).

pCDFDuet1 constructs were as follows: HSP90AA1-His_6(full length), hereafter referred to as 'HSP90'; untagged HSP90(full length); His_6-MO25a(full length); His_6-ALPHAK3/ALPK1 N-terminal domain (1–474); and His_8-CCNC(full length) in tandem with MED12-His_8(1–100); untagged MEK5/MAP2K5(S311D, T315D; full length); and untagged CK2B(full length). pET28a constructs were as follows: His_6-PDPK1(full length); His_6-PRP4/PRPF4B(519–end); GST-CAMK1A(full length); GST-CHAK2/TRPM6(1699–end); His_6-caMLCK/MYLK3(490–end); His_6-CAMKK1 (124–411); His_6-ERK7/MAPK15(full length); His_6-SUMO-ALPHAK3/ALPK1 CTD(959–end); MYO3A-His_6(1–308); ERK5/MAPK7-His_6(1–405); His_6-NIK/MAP3K14(327–673); and BMPR2-His_6(172–504). pETDuet1 constructs were as follows: His_6-CDK8(1–360, fusion with C-tail of CDK19(360–end)), His_6-CDK19(full length); His_6-AAK1(27–365); His_6-

BIKE(37–345); CK2A1-His_6(full length); CK2A2-His_6(full length); His_{10}-MBP-MEKK1/MAP3K1(1174–end); His_6-CLK1(128–end); His_8-PLK2(57–360); His_{10}-MAP3K15(631–922); His_6-SUMO-ASK1/MAP3K5(659–951); and His_6-TAO2(1–350). The pACYDuet1 construct was as follows: untagged CDC37(full length).

For enhanced expression in mammalian lines cells, the DNA sequences of His_6-GST-SBK1(full length) and Flag-His_6-WNK3(1–434) were optimized for expression in *Homo sapiens* using GeneSmart (GenScript) and synthesized as gBlocks (Integrated DNA Technologies) carrying 16 bp overhangs to facilitate in-fusion cloning into digested pCDNA3.4 (Thermo Fisher Scientific).

To generate a mammalian expression construct for the TAK1/MAP3K7, the coding sequence for this kinase (GE Healthcare Dharmacon, MHS6278-202756930) and its binding partner TAB1 (GE Healthcare Dharmacon, MHS6278-202760135) were PCR-amplified and ligated as a fusion construct (TAK1(1–303)-TAB1(451–end)) into the mammalian expression vector pLenti-X by in-fusion.

Expression constructs purchased or obtained from other laboratories or Addgene were as follows: bacterial expression constructs for GST-VRK1(full length) and GST-VRK2(full length), in pGEX-4T, were received as gifts from P. Lazo[62]. The bacterial expression construct for mouse CDKL5-His_6(1–352), in pET23a+, was received as a gift from S. Katayama[63]. Bacterial expression constructs for His_6-SUMO-PDHK1(full length), His_6-SUMO-PDHK4(full length), pGroESL (GroEL/GroES) and MBP-BCKDK(full length) were received as gifts from D. Chuang, S.-C. Tso and R. Wynn[64,65]. pProEx HTa-BRAF_16mut V600E(444–721) was a gift from M. Therrien at Université de Montréal[66]. Mammalian expression constructs for Flag-ATR(S1333A) and HA-ATRIP were provided by D. Cortez[67]. Bacterial expression constructs for DMPK1, CAMK1G, CAMK2G, PHKG2, CDKL1, GAK and lambda phosphatase were purchased from Addgene (Addgene, 1000000094)[68].

Expression and Purification from bacteria

Transformations were performed using BL21 Star cells (Thermo Fisher Scientific) unless specified otherwise. Antibiotic concentrations used were as follows: carbenicillin (100 mg l^{-1}), kanamycin (50 mg l^{-1}), spectinomycin (25 mg l^{-1}) and chloramphenicol (25 mg l^{-1} in ethanol, prepared fresh). Transformed cells were grown in 1 l Terrific Broth by shaking at 190 rpm at 37 °C until the optical density (λ = 600 nm) reached 0.7–0.8, at which point 1 mM IPTG was added to induce expression. The cells were then transferred to a refrigerated shaker and shaken at 220 rpm at 18 °C for 16–20 h. Cells were centrifuged at 6,000*g*, and the pellets were snap-frozen in liquid nitrogen and stored at −80 °C.

All of the steps for protein purification were performed at 4 °C. Cell pellets were solubilized in lysis buffer (described below) and lysed by probe sonication. The lysate was centrifuged at 20,000*g* for 1 h and the supernatant was combined with affinity purification resin, nickel NTA (Qiagen) or glutathione Sepharose (GE Health) that had been rinsed in base buffer. The supernatant-bead slurry was agitated for 30 min. Resin was washed with 1 l base buffer and eluted in 10 bed volumes of elution buffer. Eluted protein was concentrated using the Ultra Centrifugal Filter Units (Amicon), supplemented with 1 mM DTT and 25% glycerol, and snap-frozen in liquid nitrogen and stored at −80 °C.

The buffers were as follows. Standard lysis buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, HALT EDTA-free phosphatase and protease inhibitor cocktail (Life technologies), 5 mM β-mercaptoethanol and 1–3 g of lysozyme (Sigma-Aldrich). Standard base buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol (50 mM imidazole was included for purifications involving polyhistidine tags). Standard wash buffer: 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM $MgCl_2$, 2% glycerol (50 mM imidazole was included for purifications involving polyhistidine tags). Polyhistidine-tag elution buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 350 mM imidazole. GST-tag elution buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 10 mM glutathione (pH was adjusted after addition of glutathione).

CDK8 was co-purified with CCNC/MED12. CDK19 was co-purified with CCNC/MED12. CK2A1 and CK2A2 were co-purified with CK2B. ERK5

was co-expressed with MEK5DD. The kinases BRAF and NIK were co-expressed with untagged HSP90–CDC37 complex. ALPHAK3 N- and C-terminal domains were co-purified. DMPK1, CAMK1G, CAMK2G, PHKG2, CDKL1 and GAK were co-expressed with lambda phosphatase in Rosetta 2 cells (Novagen). PDHK1, PDHK4 and BCKDK were co-expressed with GroeL/GroeS and purified with the following buffers: lysis buffer (100 mM potassium phosphate pH 7.5, 10 mM l-arginine, 500 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.2% Triton X-100, lysozyme), wash buffer (50 mM potassium phosphate pH 7.5, 10 mM arginine, 500 mM NaCl, 0.1% Triton X-100, 2 mM $MgCl_2$), and elution buffer (25 mM Tris pH 7.5, 120 mM KCl, 0.02% Tween-20, 50 mM arginine, 350 mM imidazole for PDHK1 and PDHK4, 20 mM maltose for BCKDK). BCKDK was purified by its MBP tag on amylose resin (NEB). CDKL5 was expressed in BL21-codonplus(DE3)-RIL cells. KIS (full length) was purified as described previously[69].

Expression and purification from mammalian cells

Expi293F cells (Thermo Fisher Scientific) were cultured in 500 ml Expi293 Expression Medium (Thermo Fisher Scientific) in 2 l spinner flasks on a magnetic stirring platform at 100 r.c.f. at 36.8 °C under 8% CO_2. For transfection, 500 μg of expression constructs was diluted in Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific). ExpiFectamine 293 Reagent (Thermo Fisher Scientific) was diluted with Opti-MEM separately then combined with diluted plasmid DNA for 10 min at room temperature. The mixture was then transferred to the cells (3×10^6 cells per ml) and stirred. Then, 20 h after transfection, ExpiFectamine 293 Transfection Enhancer 1 and Enhancer 2 (Thermo Fisher Scientific) were added to the cells. Two days later, the cells were centrifuged at 300*g* for 5 min, snap-frozen in liquid nitrogen and stored at −80 °C (3 days after transfection).

All of the steps for protein purification were performed at 4 °C. Cell pellets were solubilized in lysis buffer and lysed by dounce homogenization (20 strokes). The lysate was centrifuged at 100,000*g* for 1 h and the supernatant was combined with affinity purification resin, nickel NTA (Qiagen), glutathione Sepharose (GE Health) or anti-Flag M2 affinity gel (Sigma-Aldrich), and agitated for 30 min (nickel and glutathione beads) or 1 h (anti-

Flag beads). Resin was washed with 1 l base buffer and eluted in 10 bed volumes of elution buffer. For elution of Flag tagged-proteins, beads were immersed in elution buffer (0.15 μg ml^{-1} 3× Flag peptide (Sigma-Aldrich)) and agitated for 1 h before elution. Eluted protein was concentrated using Ultra Centrifugal Filter Units (Amicon), supplemented with 1 mM DTT and 25% glycerol, and snap-frozen in liquid nitrogen and stored at −80 °C.

Buffers were as follows. Standard lysis buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$, 5% glycerol, 1% Triton X-100, 5 mM β-mercaptoethanol, HALT protease inhibitors. Standard base buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol. Standard wash buffer: 50 mM Tris pH 8.0, 500 mM NaCl, 2 mM $MgCl_2$, 2% glycerol.

His_6–GST-tagged SBK was purified sequentially on nickel and then glutathione resins. The buffers were as follows: the first wash buffer: 25 mM imidazole. SBK1 elution buffer for polyhistidine tag: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 250 mM imidazole. SBK1 elution buffer for GST tag: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 10 mM glutathione. Flag–TAK1–TAB1 elution buffer: 50 mM Tris pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2% glycerol, 0.15 μg ml^{-1} 3× Flag peptide.

Flag–His_6–WNK3 was purified sequentially on nickel and then anti-Flag resins. The buffers were as follows: the first wash buffer: 25 mM imidazole. Flag-tag elution buffer (chloride-free): 50 mM Tris pH 7.5, 2 mM magnesium acetate, 2% glycerol, 0.15 μg ml^{-1} 3× Flag peptide.

Flag–ATR(S1333A) (350 μl) and HA–ATRIP (150 μg) were co-transfected into Expi293 cells and incubated for one additional day after addition of enhancers (4 days after transfection). The buffers were as follows. ATR lysis buffer: 50 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol, 0.25% Tween-20, 2 mM $MgCl_2$, DTT. ATR wash buffer: 50 mM HEPES pH 7.4, 150 mM NaCl, 0.01% Brij-35, 2 mM $MgCl_2$, 5 mM ATP, DTT. ATR elution buffer: 20 mM HEPES pH 7.4, 150 mM NaCl, 0.01% Brij-35, DTT, 0.15 μg ml^{-1} 3× Flag peptide.

Eluates were concentrated to 1 ml in 100 kDa MWCO Amicon tubes and resolved using the MonoS column in a 0–1 M NaCl gradient (buffer: 25 mM Bis-Tris pH 6.9, 0.01% Brij-35 and 5 mM TCEP). A total of 1 ml of each fraction was collected. Fractions 1–4 were combined and concentrated to 1 ml using a 100 kDa MWCO filter and resolved using size-exclusion (Superose 6) in 20 mM HEPES pH 7.4, 200 mM NaCl, 0.01% Brij-35 and 5 mM TCEP. A total of 1 ml of each fraction was collected. Fractions 11–14 were verified to be pure ATR–ATRIP complex on SDS–PAGE.

SMG1–SMG9 complexes were purified from HEK293T cells as described previously[70]. RIPK1, RIPK2 and RIPK3 were purified from insect cells (Sf9) as described previously[71]. The following recombinant active kinases obtained from other laboratories. Recombinant active CDK12–CycK, CDK13–CycK and CDK9–CycT complexes were provided as gifts from M. Geyer[72,73]. Recombinant active DCAMKL1/DCLK1 and MELK were provided as gifts from N. Gray, H.-T. Huang and K. Westover, Y. Liu and W. Harshburger[74,75,76]. Recombinant active PRPK(full length)–CGI121/TPRKB(full length) complex was provided as a gift from L. Wan and F. Sicheri[77]. Recombinant active HASPIN(452–798) was provided as a gift from A. Musacchio[78]. Recombinant active YSK1 was provided as a gift from X. Luo[79]. Recombinant CK1G2 was provided as a gift from S. Knapp. A list of catalogue and lot numbers of purchased recombinant kinases is provided in Supplementary Table 1.

PSPA analysis

Recombinant kinase was added to a 384-well plate containing peptide substrate library mixtures in solution phase at 50 μM (Anaspec, AS-62017-1 and AS-62335). The reaction was initiated with the addition of 50 μM ATP (50 μCi ml^{-1} γ-^{32}P-ATP, Perkin-Elmer) and incubated for 90 min. The assay conditions for each kinase are described in Supplementary Table 1 (refs. [80,81,82,83,84]). After completion of the reaction, the solutions were spotted onto streptavidin-conjugated membranes (Promega, V2861), where the peptides tightly associated through their C-terminal biotinylation. The membranes were rinsed and then imaged using the Typhoon FLA 7000 phosphorimager (GE) to measure the extent of peptide phosphorylation. Raw data (GEL file) were quantified using ImageQuant (GE) to generate

densitometry matrices (Supplementary Table 2). For the kinase ALPHAK3, spots were normalized to the surrounding background, owing to spatial variation in background signal. PDHK1 and PDHK4 showed dual specificity for serine and tyrosine. For these kinases, we used a customized peptide substrate library devoid of tyrosine residues at randomized positions.

In total, 283 human kinase motifs, one motif from a mouse kinase orthologue (CDKL5), one motif from a rat kinase orthologue (KIS) and one motif from an arthropod *Pediculus humanus corporis* kinase orthologue (PINK1), were combined with 17 human kinase motifs that we previously published, including AKT1[85], SRPK1[35], SRPK2[35], SRPK3[35], CK1D[35], DYRK1A[86], DYRK2[86], GSK3A[86], GSK3B[86], CK1A[86], CK1E[86], CK1G1[86], CDK10[87], CDK2[88], CDK3[88], CDK18[88] and CDK7[89].

For the zero-control experiments in Extended Data Fig. 6, biotinylated peptides were synthesized containing only serine or threonine as the phospho-acceptor, where all nine surrounding positions contained degenerate mixtures of the 20 natural amino acids excluding serine, threonine, tyrosine and cysteine.

Matrix processing

The densitometry matrices were column-normalized at all positions by the sum of the 17 randomized amino acids (excluding serine, threonine and cysteine), to yield PSSMs (Supplementary Table 2). PDHK1 and PDHK4 were normalized to the 16 randomized amino acids (excluding serine, threonine, cysteine and additionally tyrosine), corresponding to the uniquely customized peptide library that profiled these kinases. The cysteine row was scaled by its median to be 1/17 (1/16 for PDHK1 and PDHK4). The serine and threonine values in each position were set to be the median of that position. The ratio of serine versus threonine phospho-acceptor favourability (S_0 and T_0, respectively) was determined by summing the values of the serine and threonine rows in the densitometry matrix (S_S and S_T, respectively), accounting for the different serine versus threonine composition of the central (1:1) and peripheral (only serine or only threonine) positions (S_{ctrl} and T_{ctrl}, respectively), and then normalizing to the

higher value among the two (S_0 and T_0, respectively, Supplementary Note 1).

Matrix clustering

The dendrogram in Fig. 2 was generated using the normalized matrices with the 20 unmodified amino acids, as well as phosphothreonine and phosphotyrosine. The linkage matrix was computed using the SciPy package in Python (v.3.7.6), using the Ward method. Results were converted to the Newick tree format and plotted using FigTree (v.1.4.4).

Substrate scoring

For scoring substrates, the values of the corresponding amino acids in the corresponding positions were multiplied and scaled by the probability of a random peptide (Supplementary Note 2).

For the percentile score of a substrate by a given kinase, we first computed the a priori score distribution of that kinase PSSM by scoring a reference Ser/Thr phosphoproteome comprising 82,735 identified sites[4] using the method discussed above (Fig. 3a). The percentile score of a kinase–substrate pair is defined as the percentile ranking of the substrate within the score distribution of each kinase[34]. This value was used when analysing all of the detected phosphorylation sites for kinase enrichment.

Kinase enrichment analysis

The single phosphorylation sites (not including multi-phosphorylated peptides) in the analysed phosphoproteomics studies were scored by all of the characterized kinases (303 Ser/Thr kinases), and their ranks in the known phosphoproteome score distribution were determined as described above. For every non-duplicate, singly phosphorylated site, kinases that ranked within the top 15 kinases for the Ser/Thr kinases were considered to be biochemically favoured kinases for that phosphorylation site. For assessing kinase motif enrichment in phosphoproteomics datasets, we compared the percentage of phosphorylation sites for which each kinase was predicted among the upregulated/downregulated (increased/decreased, respectively)

phosphorylation sites (sites with $|\log_2[\text{fold change}]|$ equal or greater than the log[fold change] threshold), versus the percentage of biochemically favoured phosphorylation sites for that kinase within the set of unregulated (unchanged) sites in this study (sites with $|\log_2[\text{fold change}]|$ less than the log[fold change] threshold). The log-transformed fold change threshold was determined to be 1.5 for all panels in Fig. 4, except for Fig. 4e, in which the threshold was set to 0.5 owing to the low range of the log[fold change] in the data. Contingency tables were corrected using Haldane correction (adding 0.5 to the cases with zero in one of the counts). Statistical significance was determined using one-sided Fisher's exact tests, and the corresponding *P* values were adjusted using the Benjamini–Hochberg procedure. Kinases that were significantly enriched (adjusted $P \leq 0.1$), or depleted ($\log_2[\text{frequency factor}] < 0$) for both upregulated and downregulated analysis were excluded from downstream analysis. Then, for every kinase, the most significant enrichment side (upregulated or downregulated) was selected on the basis of the adjusted *P* value and presented in the volcano plots.

Sequence logos

Sequence logos were made using logomaker package in Python[90]. For individual kinases, the normalized matrix was used, where the height of every letter is the ratio of its value to the median value of that position. The serine and threonine heights in the central position (position zero) were set to the ratio between their favourability. For clustered groups of kinases, the average matrix was calculated and presented as sequence logo as described above.

Comparative analyses between amino acids in the kinase domains and their substrate specificities

For Extended Data Fig. 7, kinases were sorted by their $\log_2[S_0/T_0]$ values. For the sequence logo, kinase domains of 290 available kinases were obtained from previously aligned kinase sequences[91]. The alignments to residues Met1–Leu296 in CDK2 (Protein Data Bank (PDB): 1QMZ) were obtained for each kinase, and the frequencies of amino acids in increments of 15-kinases were calculated and plotted as a sequence logo.

Known kinase–substrate pairs

Experimentally validated kinase–substrate relationships were obtained from PhosphoSitePlus (July 2021). The number of reports for each pair was determined by the sum of the in vivo and in vitro reports.

Illustrations

Experimental schema and illustrative models were generated using BioRender (https://biorender.com/). Kinome tree images were generated and modified using Coral (http://phanstiel-lab.med.unc.edu/CORAL/). Structural illustrations were generated using PyMOL. Generic kinase domains in Figs. 1 and 3 were as follows: PKAα (PDB: 1ATP). The kinase and substrate structures in Fig. 3 were as follows: ATM (PDB: 7SIC)[92] and p53 (chimera of AlphaFold AF-P04637-F1-model_v2_1 (1–95)[61] and 2ATA(96–292)[92]) (Fig. 3c), and PHKG2 (PDB: 2Y7J)[92] and PYGM (PDB: 1ABB)[92] (Fig. 3b).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated (RAW files) and analysed in this study are provided in this paper. All plasmids generated in this study have either been deposited at Addgene or available on request.

Code availability

The analytical tools used in this study and underlying code are available to the public online (https://kinase-library.phosphosite.org).

References

1. Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–E130 (2002).

2. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).

3. Hornbeck, P. V. et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.* **47**, D433–D441 (2019).

4. Ochoa, D. et al. The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373 (2020).

5. Fuhs, S. R. & Hunter, T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr. Opin. Cell Biol.* **45**, 8–16 (2017).

6. Hunter, T. Why nature chose phosphate to modify proteins. *Philos. Trans. R. Soc. B* **367**, 2513–2516 (2012).

7. Lahiry, P., Torkamani, A., Schork, N. J. & Hegele, R. A. Kinase mutations in human disease: interpreting genotype–phenotype relationships. *Nat. Rev. Genet.* **11**, 60–74 (2010).

8. Berginski, M. E. et al. The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res.* **49**, D529–D535 (2021).

9. Edwards, A. M. et al. Too many roads not taken. *Nature* **470**, 163–165 (2011).

10. Needham, E. J., Parker, B. L., Burykin, T., James, D. E. & Humphrey, S. J. Illuminating the dark phosphoproteome. *Sci. Signal.* **12**, eaau8645 (2019).

11. Lemeer, S. & Heck, A. J. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **13**, 414–420 (2009).

12. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

13. Riley, N. M. & Coon, J. J. Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* **88**, 74–94 (2016).

14. Kemp, B. E., Graves, D. J., Benjamini, E. & Krebs, E. G. Role of multiple basic residues in determining the substrate specificity of cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **252**, 4888–4894 (1977).

15. Kemp, B. E. & Pearson, R. B. Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15**, 342–346 (1990).

16. Marin, O., Meggio, F., Marchiori, F., Borin, G. & Pinna, L. A. Site specificity of casein kinase-2 (TS) from rat liver cytosol: a study with model peptide substrates. *Eur. J. Biochem.* **160**, 239–244 (1986).

17. Clark-Lewis, I., Sanghera, J. S. & Pelech, S. Definition of a consensus sequence for peptide substrate recognition by $p44^{mpk}$, the meiosis-activated myelin basic protein kinase. *J. Biol. Chem.* **266**, 15180–15184 (1991).

18. Songyang, Z. et al. Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* **4**, 973–982 (1994).

19. Hutti, J. E. et al. A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* **1**, 27–29 (2004).

20. Mok, J. et al. Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci. Signal.* **3**, ra12 (2010).

21. Pearce, L. R., Komander, D. & Alessi, D. R. The nuts and bolts of AGC protein kinases. *Nat. Rev. Mol. Cell Biol.* **11**, 9–22 (2010).

22. Kim, S.-T., Lim, D.-S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).

23. O'Neill, T. et al. Utilization of oriented peptide libraries to identify substrate motifs selected by ATM. *J. Biol. Chem.* **275**, 22719–22727 (2000).

24. Shah, N. H. et al. An electrostatic selection mechanism controls sequential kinase signaling downstream of the T cell receptor. *eLife* **5**, e20105 (2016).

25. Zhu, G. et al. Exceptional disfavor for proline at the P + 1 position among AGC and CAMK kinases establishes reciprocal specificity between them and the proline-directed kinases. *J. Biol. Chem.* **280**, 10743–10748 (2005).

26. Alexander, J. et al. Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. *Sci. Signal.* **4**, ra42 (2011).

27. Reiter, E. & Lefkowitz, R. J. GRKs and β-arrestins: roles in receptor silencing, trafficking and signaling. *Trends Endocrinol. Metab.* **17**, 159–165 (2006).

28. Moore, C. A., Milano, S. K. & Benovic, J. L. Regulation of receptor trafficking by GRKs and arrestins. *Annu. Rev. Physiol.* **69**, 451–482 (2007).

29. Bradley, D. et al. Sequence and structure-based analysis of specificity determinants in eukaryotic protein kinases. *Cell Rep.* **34**, 108602 (2021).

30. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).

31. Creixell, P. et al. Unmasking determinants of specificity in the human kinome. *Cell* **163**, 187–201 (2015).

32. Miller, M. L. et al. Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **1**, ra2 (2008).

33. Chen, C. et al. Identification of a major determinant for serine-threonine kinase phosphoacceptor specificity. *Mol. Cell* **53**, 140–147 (2014).

34. Yaffe, M. B., Leparc, G. G., Lai, J., Obata, T., Volinia, S., & Cantley, L. C. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat. Biotechnol.* **19**, 348–353 (2001).

35. Yaron, T. M. et al. Host protein kinases required for SARS-CoV-2 nucleocapsid phosphorylation and viral replication. *Sci. Signal.* **15**, eabm0808 (2022).

36. Fischer, E. H., Graves, D. J., Crittenden, E. R. S. & Krebs, E. G. Structure of the site phosphorylated in the phosphorylase *b* to *a* reaction. *J. Biol. Chem.* **234**, 1698–1704 (1959).

37. Xu, B.-e, Wilsbacher, J. L., Collisson, T. & Cobb, M. H. The N-terminal ERK-binding site of MEK1 is required for efficient feedback phosphorylation by ERK2 in vitro and ERK activation in vivo. *J. Biol. Chem.* **274**, 34029–34035 (1999).

38. Malumbres, M. et al. Cyclin-dependent kinases: a family portrait. *Nat. Cell Biol.* **11**, 1275–1276 (2009).

39. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).

40. Cohen, P. & Frame, S. The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2**, 769–776 (2001).

41. Meng, Z. et al. MAP4K family kinases act in parallel to MST1/2 to activate LATS1/2 in the Hippo pathway. *Nat. Commun.* **6**, 8357 (2015).

42. Shaywitz, A. J. & Greenberg, M. E. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu. Rev. Biochem.* **68**, 821–861 (1999).

43. Rigbolt, K. T. & Blagoev, B. Quantitative phosphoproteomics to characterize signaling networks. *Semin. Cell Dev. Biol.* **23**, 863–871

(2012).

44. Tagliabracci, V. S. et al. A single kinase generates the majority of the secreted phosphoproteome. *Cell* **161**, 1619–1632 (2015).

45. Needham, E. J. et al. Phosphoproteomics of acute cell stressors targeting exercise signaling networks reveal drug interactions regulating protein secretion. *Cell Rep.* **29**, 1524–1538 (2019).

46. Kettenbach, A. N. et al. Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. *Sci. Signal.* **4**, rs5 (2011).

47. van Vugt, M. A. et al. A mitotic phosphorylation feedback network connects Cdk1, Plk1, 53BP1, and Chk2 to inactivate the G2/M DNA damage checkpoint. *PLoS Biol.* **8**, e1000287 (2010).

48. Macůrek, L. et al. Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**, 119–123 (2008).

49. Winter, M. et al. Deciphering the acute cellular phosphoproteome response to irradiation with X-rays, protons and carbon ions. *Mol. Cell. Proteom.* **16**, 855–872 (2017).

50. Reinhardt, H. C., Aslanian, A. S., Lees, J. A. & Yaffe, M. B. p53-deficient cells rely on ATM-and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage. *Cancer Cell* **11**, 175–189 (2007).

51. Reinhardt, H. C. & Yaffe, M. B. Kinases that control the cell cycle in response to DNA damage: Chk1, Chk2, and MK2. *Curr. Opin. Cell Biol.* **21**, 245–255 (2009).

52. Xie, S. et al. Plk3 functionally links DNA damage to cell cycle arrest and apoptosis at least in part via the p53 pathway. *J. Biol. Chem.* **276**, 43305–43312 (2001).

53. Gonzalez-Hunt, C. et al. Mitochondrial DNA damage as a potential biomarker of LRRK2 kinase activity in LRRK2 Parkinson's disease.

Sci. Rep. **10**, 17293 (2020).

54. Humphrey, S. J. et al. Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab.* **17**, 1009–1020 (2013).

55. Mertins, P. et al. An integrative framework reveals signaling-to-transcription events in toll-like receptor signaling. *Cell Rep.* **19**, 2853–2866 (2017).

56. Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298**, 1911–1912 (2002).

57. Miller, C. J. & Turk, B. E. Homing in: mechanisms of substrate targeting by protein kinases. *Trends Biochem. Sci* **43**, 380–394 (2018).

58. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).

59. Joughin, B. A., Liu, C., Lauffenburger, D. A., Hogue, C. W. & Yaffe, M. B. Protein kinases display minimal interpositional dependence on substrate sequence: potential implications for the evolution of signalling networks. *Philos. Trans. R. Soc. B* **367**, 2574–2583 (2012).

60. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

61. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

62. Sanz-García, M. et al. Substrate profiling of human vaccinia-related kinases identifies coilin, a Cajal body nuclear protein, as a phosphorylation target with neurological implications. *J. Proteom.* **75**, 548–560 (2011).

63. Sekiguchi, M. et al. Identification of amphiphysin 1 as an endogenous substrate for CDKL5, a protein kinase associated with X-linked

neurodevelopmental disorder. *Arch. Biochem. Biophys.* **535**, 257–267 (2013).

64. Wynn, R. M., Davie, J. R., Cox, R. P. & Chuang, D. T. Chaperonins GroEL and GroES promote assembly of heterotetramers (*α2β2*) of mammalian mitochondrial branched-chain alpha-keto acid decarboxylase in *Escherichia coli*. *J. Biol. Chem.* **267**, 12400–12403 (1992).

65. Song, J.-L., Li, J., Huang, Y.-S. & Chuang, D. T. Encapsulation of an 86-kDa assembly intermediate inside the cavities of GroEL and its single-ring variant SR1 by GroES. *J. Biol. Chem.* **278**, 2515–2521 (2003).

66. Thevakumaran, N. et al. Crystal structure of a BRAF kinase domain monomer explains basis for allosteric regulation. *Nat. Struct. Mol. Biol.* **22**, 37–43 (2015).

67. Luzwick, J. W., Nam, E. A., Zhao, R. & Cortez, D. Mutation of serine 1333 in the ATR HEAT repeats creates a hyperactive kinase. *PLoS ONE* **9**, e99397 (2014).

68. Albanese, S. K. et al. An open library of human kinase domain constructs for automated bacterial expression. *Biochemistry* **57**, 4675–4689 (2018).

69. Manceau, V. et al. Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *FEBS J.* **273**, 577–587 (2006).

70. Melero, R. et al. Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. *Structure* **22**, 1105–1119 (2014).

71. Najjar, M. et al. Structure guided design of potent and selective ponatinib-based hybrid inhibitors for RIPK1. *Cell Rep.* **10**, 1850–1860 (2015).

72. Czudnochowski, N., Bösken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, 842 (2012).

73. Greifenberg, A. K. et al. Structural and functional analysis of the Cdk13/Cyclin K complex. *Cell Rep.* **14**, 320–331 (2016).

74. Liu, Y. et al. Chemical biology toolkit for DCLK1 reveals connection to RNA processing. *Cell Chem. Biol.* **27**, 1229–1240 (2020).

75. Ferguson, F. M. et al. Discovery of a selective inhibitor of doublecortin like kinase 1. *Nat. Chem. Biol.* **16**, 635–643 (2020).

76. Huang, H.-T. et al. MELK is not necessary for the proliferation of basal-like breast cancer cells. *eLife* **6**, e26693 (2017).

77. Wan, L. C. et al. Proteomic analysis of the human KEOPS complex identifies C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic Acids Res.* **45**, 805–817 (2017).

78. Villa, F. et al. Crystal structure of the catalytic domain of Haspin, an atypical kinase implicated in chromatin organization. *Proc. Natl Acad. Sci. USA* **106**, 20204–20209 (2009).

79. Bae, S. J., Ni, L. & Luo, X. STK25 suppresses Hippo signaling by regulating SAV1-STRIPAK antagonism. *eLife* **9**, e54863 (2020).

80. Murillo-de-Ozores, A. R., Chávez-Canales, M., de Los Heros, P., Gamba, G. & Castañeda-Bueno, M. Physiological processes modulated by the chloride-sensitive WNK-SPAK/OSR1 kinase signaling pathway and the cation-coupled chloride cotransporters. *Front. Physiol.* **11**, 585907 (2020).

81. Filippi, B. M. et al. MO25 is a master regulator of SPAK/OSR1 and MST3/MST4/YSK1 protein kinases. *EMBO J.* **30**, 1730–1741 (2011).

82. Zhou, P. et al. Alpha-kinase 1 is a cytosolic innate immune receptor for bacterial ADP-heptose. *Nature* **561**, 122–126 (2018).

83. Taipale, M. et al. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* **150**, 987–1001 (2012).

84. Klatt, F. et al. A precisely positioned MED12 activation helix stimulates CDK8 kinase activity. *Proc. Natl Acad. Sci. USA* **117**, 2894–2905 (2020).

85. Balasuriya, N. et al. Phosphorylation-dependent substrate selectivity of protein kinase B (AKT1). *J. Biol. Chem.* **295**, 8120–8134 (2020).

86. Zheng, Y. et al. Regulation of folate and methionine metabolism by multisite phosphorylation of human methylenetetrahydrofolate reductase. *Sci. Rep.* **9**, 4190 (2019).

87. Robert, T. et al. Development of a CDK10/CycM in vitro kinase screening assay and identification of first small-molecule inhibitors. *Front. Chem.* **8**, 147 (2020).

88. Ferguson, F. M. et al. Discovery of covalent CDK14 inhibitors with pan-TAIRE family specificity. *Cell Chem. Biol.* **26**, 804–817 (2019).

89. Rimel, J. K. et al. Selective inhibition of CDK7 reveals high-confidence targets and new models for TFIIH function in transcription. *Genes Dev.* **34**, 1452–1473 (2020).

90. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).

91. Modi, V. & Dunbrack Jr, R. L. A structurally validated multiple sequence alignment of 497 human protein kinase domains. *Sci. Rep.* **9**, 19790 (2019).

92. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

Acknowledgements

We thank M. Begley and T. DeFalco for helpful discussions and technical assistance; and T. Levy, S. Beausoleil, L. D'Amato, P. Lobo, M. Tran and C. Valencia for technical support. T.M.Y. thanks S. Yaron, N. Yaron, J. R. Haddad and S. Haddad for their support; and J.L.J. thanks M. Bak-Johnson for her support. This research was supported by Leukemia & Lymphoma Society FELLOW Award (to J.L.J. and L.C.C.); National Institute of Health grants P01 CA120964 (to L.C.C.), R35-CA197588 (to L.C.C.), P01-CA117969 (to L.C.C.), R35-ES028374 (to M.B.Y.), R01-CA226898 (to M.B.Y.), R01-GM104047 (to B.E.T. and M.B.Y.), U24 DK116204 (to B.E.T. and L.C.C.) and R35-GM139550 (to D.J.T.); the joint Cancer Research UK and Brain Tumour Charity funded Brain Tumour Award C42454/A28596 (to M.B.Y.); the Charles and Marjorie Holloway Foundation (to M.B.Y.); and the MIT Center for Precision Cancer Medicine. Support was also provided by the Cancer Center Support Grant P30-CA14051 and T32CA203702 (to E.R.K.) from the National Cancer Institute.

Author information

Author notes

1. These authors contributed equally: Jared L. Johnson, Tomer M. Yaron

Authors and Affiliations

1. Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA

   Jared L. Johnson, Tomer M. Yaron, Emily M. Huntsman, Alexander Kerelsky, Junho Song, Amit Regev, Ting-Yu Lin, Katarina Liberatore, Daniel M. Cizin, Benjamin M. Cohen, Yilun Ma, Edward R. Kastenhuber, Marcus D. Goncalves, John Blenis & Lewis C. Cantley

2. Department of Medicine, Weill Cornell Medicine, New York, NY, USA

   Jared L. Johnson, Tomer M. Yaron, Emily M. Huntsman, Alexander Kerelsky, Junho Song, Amit Regev, Ting-Yu Lin, Katarina Liberatore, Daniel M. Cizin, Benjamin M. Cohen, Yilun Ma, Edward R. Kastenhuber & Lewis C. Cantley

3. Englander Institute for Precision Medicine, Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA

Tomer M. Yaron, Alexander Kerelsky & Olivier Elemento

4. Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA

Tomer M. Yaron & Olivier Elemento

5. Tri-Institutional PhD Program in Computational Biology & Medicine, Weill Cornell Medicine, Memorial Sloan Kettering Cancer Center and The Rockefeller University, New York, NY, USA

Tomer M. Yaron

6. Weill Cornell Graduate School of Medical Sciences, Cell and Developmental Biology Program, New York, NY, USA

Ting-Yu Lin

7. Department of Medicine, Division of Hematology/Oncology, Columbia University Irving Medical Center, New York, NY, USA

Neil Vasan

8. Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA

Neil Vasan

9. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Konstantin Krismer

10. Center for Precision Cancer Medicine, Koch Institute for Integrative Cancer Biology, Departments of Biology and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Konstantin Krismer, Bert van de Kooij, Anne E. van Vlimmeren & Michael B. Yaffe

11. Department of Pharmacology, Yale School of Medicine, New Haven, CT, USA

Jaylissa Torres Robles & Benjamin E. Turk

12. Department of Chemistry, Yale University, New Haven, CT, USA

Jaylissa Torres Robles

13. Institute of Genetics, Technische Universität Braunschweig, Braunschweig, Germany

Nicole Andrée-Busch & Norbert F. Käufer

14. Department of Pharmacology, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ, USA

Maxim V. Dorovkov & Alexey G. Ryazanov

15. Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA

Yuichiro Takagi

16. Division of Endocrinology, Weill Cornell Medicine, New York, NY, USA

Marcus D. Goncalves

17. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Benjamin D. Hopkins

18. Department of Biochemistry, University of Colorado, Boulder, CO, USA

Dylan J. Taatjes

19. SABNP, Univ Evry, INSERM U1204, Université Paris-Saclay, Evry, France

Alexandre Maucuer

20. Department of Investigative Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho, Japan

Akio Yamashita

21. Department of Developmental, Molecular and Chemical Biology, Tufts University School of Medicine, Boston, MA, USA

Alexei Degterev

22. Department Of Bioinformatics, Cell Signaling Technology, Danvers, MA, USA

Mohamed Uduman, Jingyi Lu, Sean D. Landry, Bin Zhang, Ian Cossentino & Peter V. Hornbeck

23. Rewire Tx, Humboldt-Universität zu Berlin, Berlin, Germany

Rune Linding

24. Department of Pharmacology, Weill Cornell Medicine, New York, NY, USA

John Blenis

25. Department of Biochemistry, Weill Cornell Medicine, New York, NY, USA

John Blenis

26. Divisions of Acute Care Surgery, Trauma, and Surgical Critical Care, and Surgical Oncology, Department of Surgery, Beth Israel Deaconess

Medical Center, Harvard Medical School, Boston, MA, USA

Michael B. Yaffe

27. Surgical Oncology Program, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Michael B. Yaffe

Contributions

J.L.J., T.M.Y., L.C.C., M.B.Y. and B.E.T. conceived the project, designed experiments and analysed the data. J.L.J. and T.M.Y. generated figures. J.L.J. performed the PSPA experiments. T.M.Y. led the computational analyses. T.M.Y., E.M.H., A.K., D.M.C., B.M.C., K.K., M.U., J.L., S.D.L., B.Z. and I.C. performed computational analyses. J.L.J., J.S., A.R., T.-Y.L., N.V., K.L., Y.M., A.D., A.Y. and A.M. generated recombinant proteins. J.L.J., T.M.Y., B.E.T., J.T.R., M.B.Y. and L.C.C. performed structural modelling. P.V.H., D.J.T., Y.T., N.A.-B., N.F.K., B.v.d.K., A.E.v.V., M.V.D., A.G.R., E.R.K., M.D.G., B.D.H., O.E., R.L. and J.B. contributed data and participated in discussions. J.L.J., T.M.Y., M.B.Y., B.E.T. and L.C.C. wrote and edited the manuscript with input from all of the authors.

Corresponding authors

Correspondence to Benjamin E. Turk, Michael B. Yaffe or Lewis C. Cantley.

Ethics declarations

Competing interests

L.C.C. is a founder and member of the board of directors of Agios Pharmaceuticals and is a founder and receives research support from Petra Pharmaceuticals; is listed as an inventor on a patent (WO2019232403A1, Weill Cornell Medicine) for combination therapy for PI3K-associated disease or disorder, and the identification of therapeutic interventions to

improve response to PI3K inhibitors for cancer treatment; is a co-founder and shareholder in Faeth Therapeutics; has equity in and consults for Cell Signaling Technologies, Volastra, Larkspur and 1 Base Pharmaceuticals; and consults for Loxo-Lilly. M.B.Y receives research support from Cardiff Oncology. T.M.Y. is a co-founder and stockholder and is on the board of directors of DESTROKE, an early-stage start-up developing mobile technology for automated clinical stroke detection. J.L.J has received consulting fees from Scorpion Therapeutics and Volastra Therapeutics. O.E. is a founder and equity holder of Volastra Therapeutics and OneThree Biotech; is a member of the scientific advisory board of Owkin, Freenome, Genetic Intelligence, Acuamark and Champions Oncology; and receives research support from Eli Lilly, Janssen and Sanofi. D.J.T. is a member of the scientific advisory board at Dewpoint Therapeutics. A.D. is an equity holder of Denali Therapeutics; and receives research support from Interline Therapeutics. N.V. reports consulting activities for Novartis and is on the scientific advisory board of Heligenics. M.D.G. is a co-founder and shareholder of Faeth Therapeutics, which is developing dietary and pharmacological therapies for cancer; and has received speaking and/or consulting fees from Pfizer, Novartis, Scorpion Therapeutics and Faeth Therapeutics.

Peer review

Peer review information

Nature thanks Tony Hunter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Representation of phosphorylation site motifs in the human serine and threonine phosphoproteome.

Venn diagram representation of the percentages of three prominent Ser/Thr kinase motif features, pertaining to Clusters 1, 2, and 3 in Fig. 2, across 82,735 human serine and threonine phosphorylation sites confidently identified in mass spectrometry experiments[4]. The phosphorylated residues in the logos are represented as S/T.

Extended Data Fig. 2 Subcategorization of the basophilic kinases of Cluster 1.

Subcategorization of Cluster 1 from Fig. 2 into 11 motif classes.

Extended Data Fig. 3 Subcategorization of the proline-directed kinases of Cluster 2.

Subcategorization of Cluster 2 from Fig. 2 into 5 motif classes.

Extended Data Fig. 4 Subcategorization of the acidophilic kinases of Cluster 3.

Subcategorization of Cluster 3 from Fig. 2 into 8 motif classes.

Extended Data Fig. 5 Structural models of kinase–substrate complexes.

a, Synthetic peptide from its complex with PAK4 (PDB: 2Q0N) modelled onto WNK3 (PDB: 5O26). Dotted circle highlights a shallow hydrophobic pocket accommodating a +3 Phe residue. **b**, GSK3 peptide from its complex with AKT2 (PDB: 1O6L) modelled onto CAMKK2 (PDB: 2ZV2). Circle indicates a hydrophobic pocket that could accommodate a −2 aliphatic residue. **c**, Monophosphorylated peptide from p63 bound to CK1δ (PDB: 6RU6) modelled onto GRK2 (PDB: 1YM7). Circle shows positive surface potential in the vicinity of the −2 and −3 pSer residues. **d**, p63 peptide bound to CK1δ (PDB: 6RU8) was modelled onto YANK1 (PDB: 4FR4) showing

potential binding sites for −3 and +2 phosphorylated residues. Surface electrostatics are represented with Coulombic potential values were computed in ChimeraX and represented by scale bars (kcal/mol·e).

Extended Data Fig. 6 Profiling phospho-acceptor specificity.

a, *in vitro* phosphorylation assays with recombinant kinases and substrate peptides containing either serine or threonine phospho-acceptors. Results shown for 208 recombinant Ser/Thr kinases. **b**, Correlation plot of the experimental results in (**a**) with the position sum approach applied in this study to score Ser/Thr phospho-acceptor preference.

Extended Data Fig. 7 Global analysis of the relationship between the DFG+1 amino acid and preference for the serine versus threonine phospho-acceptor.

a, Bottom, relative preferences for Ser or Thr phospho-acceptor residues for each kinase, arranged in order of decreasing Ser/Thr selectivity. Top, frequency of amino acids at the DGF+1 positions of corresponding kinases (bin size: 15 kinases). **b**, Structural illustration of the proximity between the DFG+1 residue and substrate phospho-acceptor residue, shown with the AKT1 kinase domain bound to substrate (GSK3β) peptide (pdb 1O6K).

Extended Data Fig. 8 Global performance analysis of substrate percentile scores for their literature-annotated kinases.

a, Percentile-score distributions of substrates for their literature-annotated kinases (AUCDF= area under the cumulative distribution function). **b**, Percentile-score of literature-annotated kinase–substrate pairs as a function of number of reports. Higher number of reports correlates with more favourable percentile-scores between the reported kinase and its substrate. n = 9,073, n = 3,945, n = 544, n = 224, and n = 201 for kinase–substrate relationships with 1, 2, 3, 4, and 5 or more reports, respectively. Statistical analyses were performed using Double-sided Mann-Whitney U-test. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extend from the box maxima or minima to the largest or smallest

value no further than 1.5 x interquartile range. (ns $p > 0.05$, * $p \leq 0.05$, ** $p \leq 10^{-3}$, *** $p \leq 10^{-4}$, **** $p \leq 10^{-5}$).

Extended Data Fig. 9 Global performance analysis of kinase ranks for their literature-annotated substrates.

a, Rank distributions of kinases for their literature-annotated substrates (AUCDF= area under the cumulative distribution function). **b**, Rank of the literature-annotated kinase–substrate pairs, as a function of number of reports. Higher number of reports correlates with more favourable ranking of reported kinase for its substrate. n = 9,073, n = 3,945, n = 544, n = 224, and n = 201 for kinase–substrate relationships with 1, 2, 3, 4, and 5 or more reports, respectively. Statistical analyses were performed using Double-sided Mann-Whitney U-test. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extend from the box maxima or minima to the largest or smallest value no further than 1.5 x interquartile range. (ns $p > 0.05$, * $p \leq 0.05$, ** $p \leq 10^{-3}$, *** $p \leq 10^{-4}$, **** $p \leq 10^{-5}$).

Extended Data Fig. 10 Motif-based scoring results for phosphorylation events facilitated by subcellular localization or docking.

a, Illustration of the mitochondrial-localized regulation of the pyruvate dehydrogenase complex through phosphorylation by the PDHKs. Scoring results for PDHA1 Ser293, highlighting the PDHK family kinases. **b**, Illustration of docking-driven phosphorylation of ERK1/2 by MEK1/2. Scoring results for ERK1 Thr202 / ERK2 Thr203 (identical sequences), highlighting MEK1 and MEK2.

Extended Data Fig. 11 Scoring comparison of CDK subfamilies.

Illustration of the phosphoregulation of RNA Polymerase II (POLR2A) CTD and Retinoblastoma protein (Rb) by their respective canonical CDKs, the transcriptional CDKs (purple) and the cell cycle progression CDKs (green) (left). Links between kinases and substrates correspond to favourable scores between motifs and phosphorylation sites (right).

Supplementary information

Supplementary Information

Supplementary Figs. 1 and 2 and Supplementary Notes 1 and 2.

Reporting Summary

Supplementary Table 1

Profiling Ser/Thr kinase substrate specificity. Experimental details for obtaining and profiling the recombinant Ser/Thr kinases in this study.

Supplementary Table 2

PSPA data and PSSMs. Raw densitometries obtained from PSPA experiments and their normalized forms.

Supplementary Table 3

Annotation of the human Ser/Thr phosphoproteome. A total of 89,752 experimentally identified Ser and Thr phosphorylation sites were scored by 303 Ser/Thr kinase PSSMs. The table enables one to sort substrates by percentile scores or ranks for given kinases or by promiscuity indices (the number of kinases scoring above the 90th percentile) or median percentile scores (Fig. 3d).

Supplementary Table 4

Motif-enrichment analysis with ATM. Serine and threonine phosphorylation sites upregulated after treatment with ionizing radiation (Fig. 4e), where ATM was ranked within the top 15 out of 303 predicted kinases.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Johnson, J.L., Yaron, T.M., Huntsman, E.M. *et al.* An atlas of substrate specificities for the human serine/threonine kinome. *Nature* **613**, 759–766 (2023). https://doi.org/10.1038/s41586-022-05575-3

- Received: 01 May 2022
- Accepted: 17 November 2022
- Published: 11 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05575-3

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

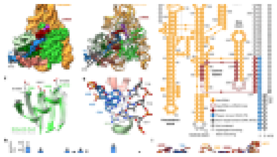
Targets mapped for almost all human kinase enzymes

- Sean J. Humphrey
- Elise J. Needham

Nature News & Views 11 Jan 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05575-3

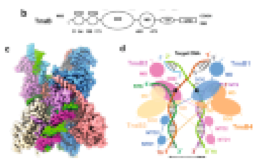
| Section menu | Main menu |



Target site selection and remodelling by type V CRISPR-transposon systems

10 November 2021

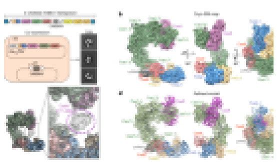
Irma Querques, Michael Schmitz, … Martin Jinek



Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon

02 October 2022

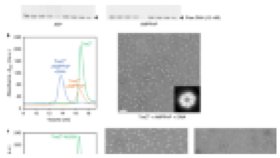
Francisco Tenjo-Castaño, Nicholas Sofos, … Guillermo Montoya



Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system

18 December 2019

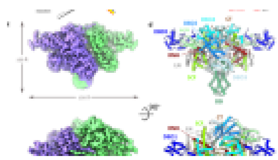
Tyler S. Halpin-Healy, Sanne E. Klompe, … Israel S. Fernández



[Structural basis for DNA targeting by the Tn7 transposon](#)

16 February 2022

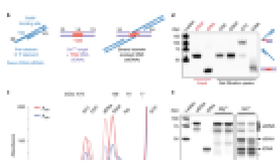
Yao Shen, Josue Gomez-Blanco, … Alba Guarné



[Structural insight into Tn3 family transposition mechanism](#)

18 October 2022

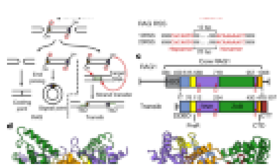
Alexander V. Shkumatov, Nicolas Aryanpour, … Rouslan G. Efremov



[Structure of a P element transposase–DNA complex reveals unusual DNA structures and GTP-DNA contacts](#)

28 October 2019

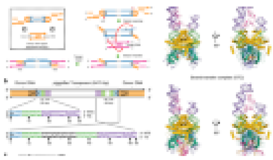
George E. Ghanim, Elizabeth H. Kellogg, … Donald C. Rio



Structures of a RAG-like transposase during cut-and-paste transposition

13 November 2019

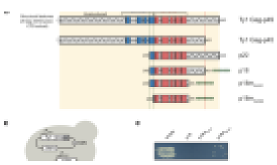
Chang Liu, Yang Yang & David G. Schatz



Structural basis of seamless excision and specific targeting by piggyBac transposase

10 July 2020

Qiujia Chen, Wentian Luo, … Fred Dyda



Structure of a Ty1 restriction factor reveals the molecular basis of transposition copy number control

22 September 2021

Matthew A. Cottee, Sean L. Beckwith, … Ian A. Taylor

- Article
- Open Access
- Published: 28 November 2022

Structures of the holo CRISPR RNA-guided transposon integration complex

- Jung-Un Park,
- Amy Wei-Lun Tsai,
- Alexandrea N. Rizo,
- Vinh H. Truong,
- Tristan X. Wellner,
- Richard D. Schargel &
- …
- Elizabeth H. Kellogg

Show authors

Nature **volume 613**, pages 775–782 (2023)

- 7605 Accesses
- 1 Citations
- 35 Altmetric
- Metrics details

Abstract

CRISPR-associated transposons (CAST) are programmable mobile genetic elements that insert large DNA cargos using an RNA-guided mechanism[1,2,3]. CAST elements contain multiple conserved proteins: a CRISPR effector (Cas12k or Cascade), a AAA+ regulator (TnsC), a transposase (TnsA–TnsB) and a target-site-associated factor (TniQ). These components are thought to cooperatively integrate DNA via formation of a multisubunit transposition integration complex (transpososome). Here we

reconstituted the approximately 1 MDa type V-K CAST transpososome from *Scytonema hofmannii* (*Sh*CAST) and determined its structure using single-particle cryo-electon microscopy. The architecture of this transpososome reveals modular association between the components. Cas12k forms a complex with ribosomal subunit S15 and TniQ, stabilizing formation of a full R-loop. TnsC has dedicated interaction interfaces with TniQ and TnsB. Of note, we observe TnsC–TnsB interactions at the C-terminal face of TnsC, which contribute to the stimulation of ATPase activity. Although the TnsC oligomeric assembly deviates slightly from the helical configuration found in isolation, the TnsC-bound target DNA conformation differs markedly in the transpososome. As a consequence, TnsC makes new protein–DNA interactions throughout the transpososome that are important for transposition activity. Finally, we identify two distinct transpososome populations that differ in their DNA contacts near TniQ. This suggests that associations with the CRISPR effector can be flexible. This *Sh*CAST transpososome structure enhances our understanding of CAST transposition systems and suggests ways to improve CAST transposition for precision genome-editing applications.

Main

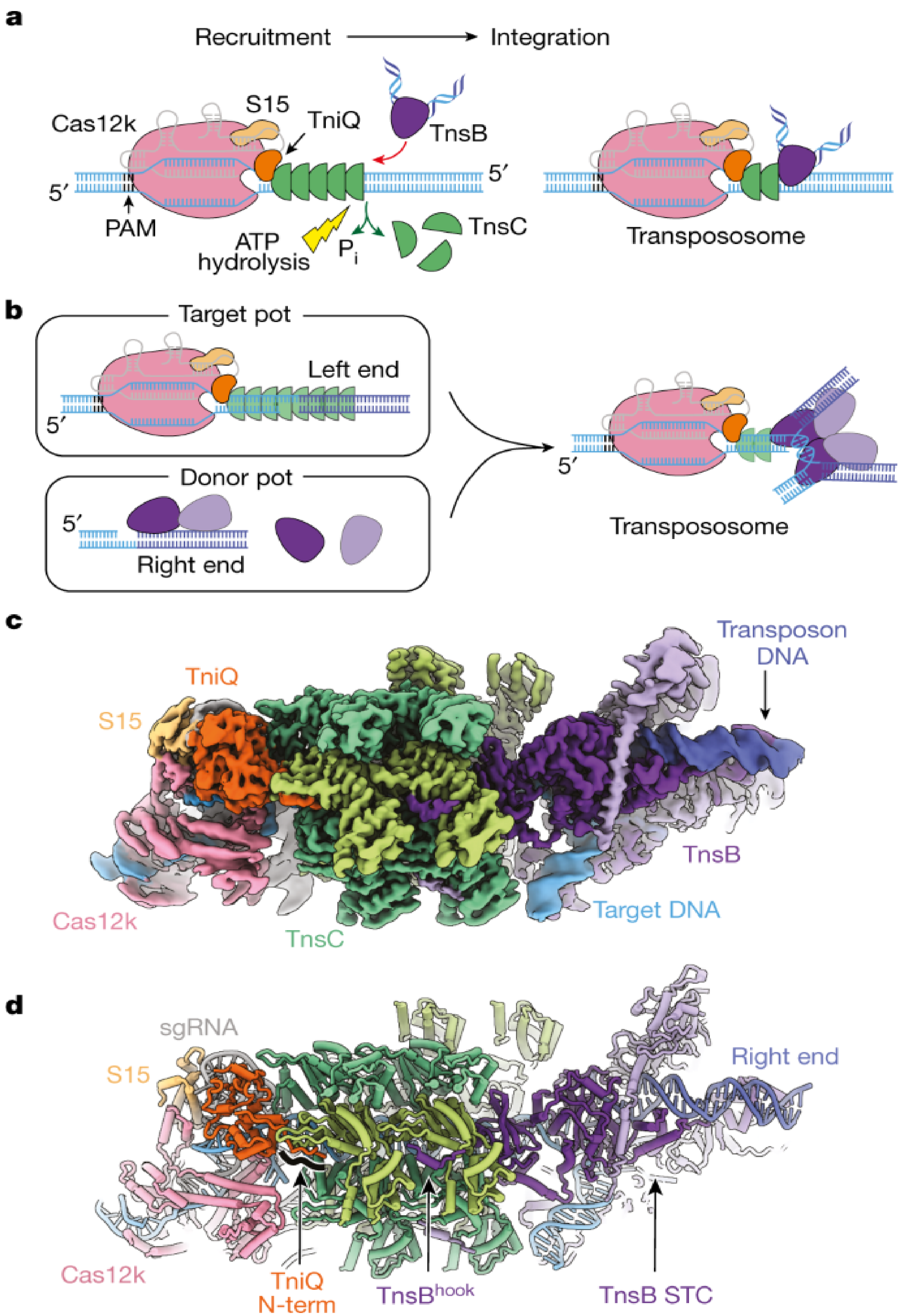
CAST are Tn7-like transposons[1] that programmably integrate large DNA cargos at genomic locations dictated by a guide RNA sequence[2,3]. Multiple independent acquisition events[1,4] created the distinct CAST subfamilies characterized to date: I-F3[3,5], I-B[6] and V-K CAST[2]. One of the most remarkable aspects of CAST function is the precision of the programmable insertions: DNA cargo is inserted in a single orientation, with defined spacing from the protospacer adjacent motif (PAM) and within a narrow window (5–10 bp). These features appear to be generally conserved across CAST families[3,5,6,7], and are probably a consequence of the transpososome architecture (that is, the nucleoprotein integration complex containing all CAST components).

Across all CAST elements[4], four or more proteins make up the CAST transposition machinery: the CRISPR effector (Cas12k or Cascade), TniQ, TnsC and the transposase (TnsB or TnsA–TnsB). In all Tn7-like transposition systems, a specialized CRISPR effector or DNA-binding

protein (TnsD in the case of the prototypic Tn7)[8] is hypothesized to recruit core transposition proteins to the target site[9]. In CAST elements, the TniQ protein (which is related to TnsD) associates with the target site via interactions with the CRISPR effector[10]. The AAA+ regulator TnsC serves as a molecular matchmaker by interacting with both target-site-binding proteins and the transposase[11,12]. Very little structural information exists to explain how TnsC bridges the components in the hypothesized transpososome assembly. Furthermore, how the core transposition machinery evolved to adopt different targeting strategies[6,13,14] remains an open question.

Owing to its simplicity and robust in vitro activity[2], ShCAST is an attractive system for understanding the general mechanisms used by CAST elements. Although the structure of each *Sh*CAST component has been individually characterized[15,16,17,18], the spatial associations across the entire transposition recruitment process remain unknown. We reconstituted the *Sh*CAST transpososome by mimicking the natural configuration found in programmed integration (Fig. 1a). Using single-particle cryo-electron microscopy (cryo-EM), we obtained a high-resolution structure (3.5 Å resolution) that was sufficient to accurately resolve the interactions between all components within the 961 kDa nucleoprotein complex.

Fig. 1: Cryo-EM structure of *Sh*CAST transpososome.


a
Recruitment
Integration
Cas12k
S15
TniQ
TnsB
5′
5′
PAM
ATP
hydrolysis
P$_i$
TnsC
Transpososome
b
Target pot
Left end
5′
Donor pot
5′
Right end
5′
Transpososome
c
Transposon
DNA
TniQ
S15
TnsB
Target DNA
Cas12k
TnsC
d
sgRNA
S15
Right end
Cas12k
TniQ
N-term
TnsB^{hook}
TnsB STC

a, Mechanistic model of *Sh*CAST recruitment (left) and integration (right). TnsC (green semicircle) associates with target-site proteins: Cas12k (pink), TniQ (orange) and S15 (tan). The PAM (black) defines the beginning of the protospacer, where Cas12k and single guide RNA (sgRNA) (grey) bind. TnsB (purple) is recruited to the target site through TnsC (red arrow). ATP hydrolysis (yellow lightning bolt) is stimulated by TnsB, resulting in the release of phosphate (P_i) and disassembly of TnsC (green arrows). Upon integration, a nucleoprotein complex containing all CAST components forms at the target site (right). **b**, Schematic of transpososome sample preparation. Left, the target pot and the donor pot were prepared independently to mimic the process of the RNA-guided transposition. DNA substrates for both the target pot and donor pot have target DNA (light blue) and transposon-end (dark blue) regions that are connected by 5 bp of single-stranded DNA, designed to form a strand-transfer complex (STC) (Methods). The target pot and donor pot include target-site-associated proteins (Cas12k, S15, TniQ and TnsC) and TnsB, respectively, in addition to the corresponding DNA substrate. The transpososome (right) is reconstituted by combining the target and donor pots (Methods). **c**, The 3.5 Å-resolution cryo-EM reconstruction of the *Sh*CAST transpososome, filtered according to local resolution. Each component in the complex is coloured as in **a**. Different shades (light or dark green and light or dark purple) indicate different subunits of the protein of the same colour. Target DNA is shown in light blue and transposon DNA is in dark blue. **d**, The atomic model of the *Sh*CAST transpososome. TniQ N terminus comprises residues 1–10. Right and left refer to the ends of the transposon DNA.

We found that the transpososome architecture promotes modular association across CAST components, and TnsC has dedicated faces of interaction with TniQ and TnsB, consistent with models of Tn7 transposition[12,19]. Cas12k, TniQ and S15 stabilize R-loop formation, but TniQ and TnsC form the primary connection between the CRISPR effector and transposase. We also observed TnsB–TnsC interactions that are important for TnsB-promoted ATP hydrolysis. Notably, in the context of the transpososome, TnsC protomers make functionally important interactions with the target DNA that are distinct from interactions found in helical TnsC[18]. Finally, we found that the transpososomes assemble with a slightly different number of TnsC protomers (±1 protomer), hinting at the molecular basis of the intrinsic

variability of *Sh*CAST insertions. The structural insight gained from this transpososome structure contributes to the understanding of CAST transposition and serves to identify avenues for protein engineering.

DNA design and transpososome reconstitution

*Sh*CAST insertions occur within a 61- to 66-base pair (bp) window downstream from the PAM in a single orientation[2,18]. Because transposition is driven primarily via protein–DNA interactions, the transpososome is the most stable structure in the transposition pathway. This principle has been applied successfully in the past to stabilize STC structures of other integrases and transposases, such as PFV[20,21]. Thus, we designed a strand-transfer substrate (Extended Data Fig. 1a) that contains binding sites for all *Sh*CAST components[2]. Our designed DNA substrate contains transposon DNA up to the first two internal TnsB binding sites from the right and left ends (Extended Data Fig. 1b), identical to previous studies[17,22]. The first 30 base pairs have identical TnsB binding sites on either end[17], in contrast to the subsequent internal TnsB binding sites that are irregularly spaced in the right transposon end compared with the left transposon end. Therefore, the transposon sequences that we included in our designed substrate most probably do not contribute to the remarkable ability of *Sh*CAST to discriminate the insertion orientation[2]. The 'target pot' from the transpososome reconstitution procedure (Fig. 1b) is similar to the procedure used to reconstitute a *Sh*CAST subcomplex referred to as the 'recruitment complex'[23], which contains all the components except TnsB. We also based our reconstitution procedure on established methods for monitoring transposition in vitro[18,24] and reasoned that this procedure would mimic the recruitment and integration process (Fig. 1b). Incubation of target DNA with target-site-binding proteins (Cas12k, TniQ, S15 and TnsC) and ATP resulted in a heterogeneous assembly, comprising different TnsC filament lengths (as assessed by negative-stain electron microscopy; Extended Data Fig. 1c). Subsequent incubation with transposon-DNA-bound TnsB, followed by enrichment for DNA-bound complexes resulted in the successful reconstitution of complete transpososome particles (Fig. 1b and Extended Data Fig. 1d), which we analysed by high-resolution cryo-EM imaging and image analysis (Methods and Extended Data Fig. 2).

Overall transpososome architecture

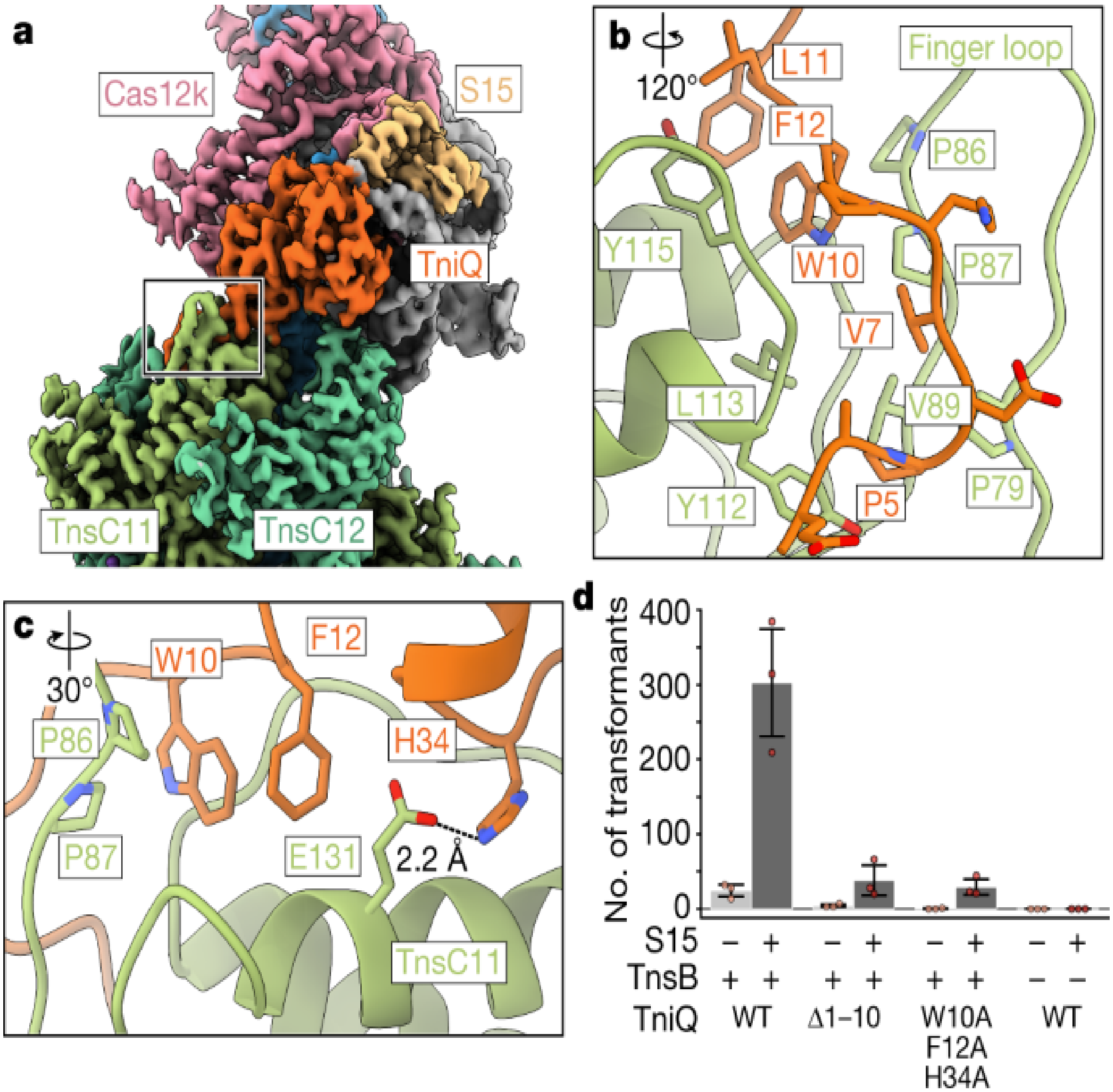
A 3.5 Å cryo-EM reconstruction (Extended Data Table 1) of the reconstituted *Sh*CAST transpososome reveals an assembly composed of one Cas12k subunit, one TniQ subunit, one ribosomal S15 subunit, four subunits of TnsB, and two full turns of TnsC with six subunits per turn (Fig. 1c). The defined oligomeric assembly of TnsC is important, because the *Sh*CAST recruitment complex (containing all components except TnsB) consists of heterogeneous assemblies of TnsC, distinguished by the direction of TnsC filaments bound to DNA[23] and the number of turns of TnsC. In contrast to the heterogeneity of the recruitment complex (captured in the target pot reconstitution, Extended Data Fig. 1c), transpososome particles reveal uniform TnsC directionality (Fig. 1c,d), in which the N-terminal face of TnsC interacts with TniQ at the target site and the C-terminal face of TnsC interacts with TnsB (Extended Data Fig. 3). In addition, TnsC filaments in transpososome particles are of uniform stoichiometry, consisting of two turns of TnsC. TnsC is disassembled by TnsB during recruitment[18] (Fig. 1a). Therefore, the homogeneity of the TnsC oligomeric assembly that we observe in our transpososome particles suggests that (1) TnsC protomers lacking productive interactions with target-site proteins are presumably disassembled by TnsB, and (2) TnsB disassembles TnsC filaments until a certain point, reflected in the uniform stoichiometry of TnsC oligomers observed in our transpososome reconstitution. Therefore, the interactions between TnsC and target-site-associated proteins (Cas12k, TniQ and S15) are stabilized against further disassembly by TnsB, probably owing to interactions at the target site (Fig. 1a, right).

Transposition requires TniQ–TnsC contacts

Owing to slight flexibility in the DNA substrate, the distal ends of the transpososome exhibit lower local resolution in our structure (5–7 Å; Extended Data Fig. 2e). As expected, local refinement focused on the Cas12k-proximal region improved the quality of the reconstruction (Supplementary Fig. 1a,b). Consistent with the functional importance of TniQ in *Sh*CAST transposition[16], TniQ is located at the PAM-distal end of the R-loop close to Cas12k and S15 (Extended Data Fig. 4a). TniQ primarily

interacts with TnsC and RNA, consistent with the productive recruitment complex[23]. S15 is nestled between the REC2 domain of Cas12k and the PAM-distal sgRNA–DNA heteroduplex (Extended Data Fig. 4d). The rooftop loop of the sgRNA is flanked on either side by S15 and TniQ, respectively. TniQ bridges the two TnsC protomers closest to Cas12k (Fig. 2a and Supplementary Fig. 2), however the TniQ–TnsC12 interface is much smaller than the TniQ–TnsC11 interface (325 $Å^2$ versus 915 $Å^2$; Methods). More specifically, the N-terminal tail of TniQ forms structured interactions with the finger loop of TnsC11 and is generally hydrophobic in nature (Fig. 2b). From the structure, we identified hydrophobic residues W10 and F12, which are buried into a hydrophobic cleft in TnsC11 formed by the finger loop (Fig. 2b), as important residues. H34 forms hydrogen-bonding interactions with the acidic residue E131 on the N-terminal face of TnsC11 (Fig. 2c). Truncation of the first ten residues or mutation of the hydrophobic residues W10, F12 and H34 results in a near-complete loss of transposition activity (Fig. 2d), consistent with the essential role suggested by our structural observations. Consistent with the importance of S15 for formation of the productive recruitment complex[23], we find that S15 generally improves transposition activity (Fig. 2d). However, the addition of S15 does not change our conclusions regarding the importance of the TniQ residues (Fig. 2d). Cas12k–TnsC interactions are completely absent (Extended Data Fig. 4c), and 3D variability analysis suggests that Cas12k enables flexible linkage to the other *Sh*CAST components (Supplementary Video 1). Together, these observations collectively suggest that the TniQ–TnsC interactions that we observe are the primary connections required to direct insertions to a guide RNA-directed target site. This supports the idea that engineering new associations with target-site recognition modules (that is, novel non-CAST CRISPR effectors) can be achieved by focusing on TniQ.

Fig. 2: TniQ–TnsC interactions are crucial for *Sh*CAST transposition.



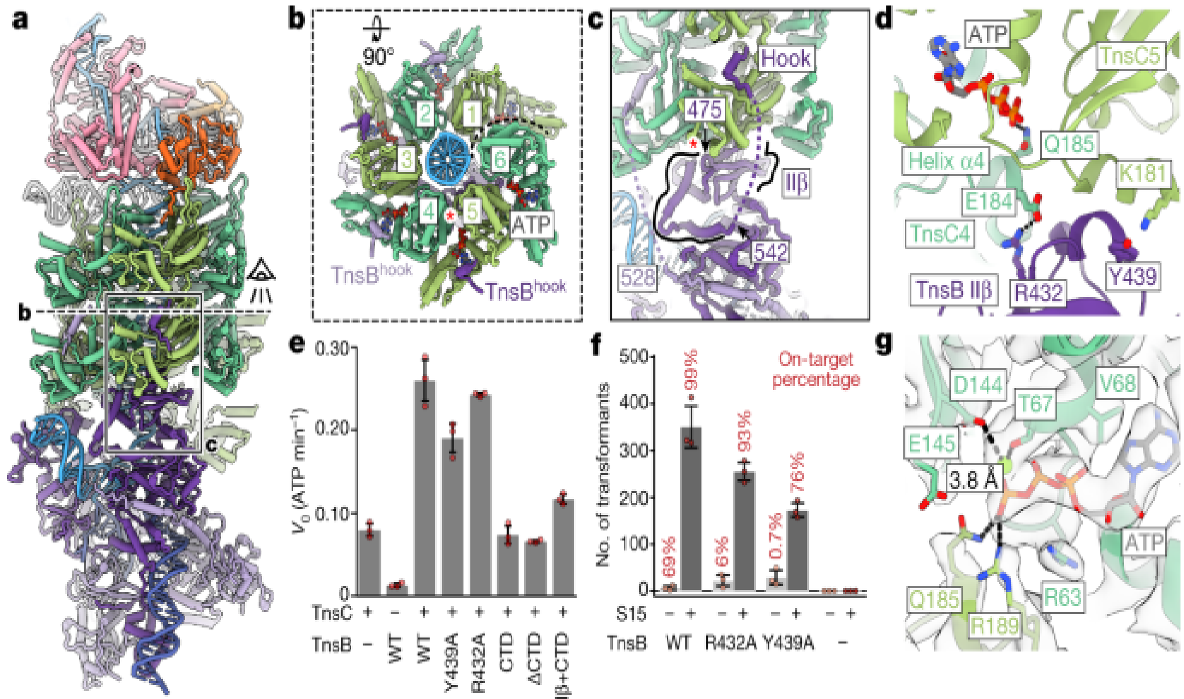
a, The composite cryo-EM map was created using two separate reconstructions from the local refinements (Methods). The composite map is coloured as in Fig.1. TnsC protomers are labelled according to their position in the transpososome (TnsC1 is the closest TnsC protomer to TnsB). The outline indicates the TniQ–TnsC interface. **b**,**c**, The atomic model of interactions at the TniQ–TnsC interface rotated 120° around the vertical axis (**b**) and 30° around the horizontal (**c**) with respect to the outlined region in **a**. **b**, The N-terminal tail of TniQ is shown—as sticks—interacting with a hydrophobic cleft created by the finger loop of TnsC. **c**, Hydrogen-bonding interactions are shown as dashed black lines; distance between the donor and acceptor atoms is indicated. **d**, TniQ mutations eliminate in vitro

transposition activity, consistent with the interactions observed in our transpososome structure. WT, wild type. In vitro transposition activity was monitored by transforming the reaction product into competent cells, and counting the number of transformants after plating on an antibiotic-containing plate (Methods). Data are mean ± s.d. ($n = 3$ biological triplicates). Raw data points are shown in red.

TnsB forms structured interactions with TnsC

We next focused on the interactions between TnsB and TnsC in the transpososome structure (Fig. 3a). TnsB belongs to the DDE/D transposase family, and bears significant similarities to MuA from bacteriophage Mu[17,22,25,26]. TnsB contains multiple functionally distinct domains dedicated to DNA-binding, catalytic activity and interactions with TnsC[17] (Supplementary Fig. 3). The previously determined tetrameric TnsB STC structure[17] (PDB: 7SVW) docks well into the transpososome cryo-EM map, requiring only minimal modifications throughout (1.54 Å Cα root mean squared deviation (r.m.s.d.), Supplementary Fig. 4). Full-length TnsB has a 52-residue-long flexible linker (residues 518–569), that connects the STC to the C-terminal 'hook' ($TnsB^{hook}$, residues 570–584) (Supplementary Fig. 3). $TnsB^{hook}$ serves as the point of contact with the TnsC filament body and is critically important for CAST transposition[17]. We identified well-defined density corresponding to the $TnsB^{hook}$ at 4 of the 6 TnsC protomers closest to TnsB (TnsC protomers 2–5) (Fig. 3b and Extended Data Fig. 5). After docking in TnsB STC structure[17], we found additional unassigned density corresponding to TnsB residues 475–542. These residues, which were previously missing in STC and predicted to be disordered by DISOPRED3[27] (Supplementary Fig. 3), form a helix–turn–helix motif (Fig. 3c). Therefore, although TnsB–TnsC associations may be flexible in principle, we see very well-defined TnsB–TnsC interactions in the transpososome, with all four TnsB protomers engaged at precise locations on the TnsB-proximal TnsC hexamer.

Fig. 3: TnsB–TnsC interactions are well defined and contribute to stimulating the ATPase activity of TnsC.



a, An overview of TnsB–TnsC interactions in the transpososome. **b**, The N-terminal face of the TnsC hexamer closest to TnsB, as indicated in **a**. TnsC protomers are labelled according to their position (1 indicates the TnsC protomer closest to TnsB). The dashed line indicates where the hexamer ends. TnsB^{hook} peptides bind at TnsC protomers 2–5 and are coloured according to nearest TnsB protomer. The red asterisk indicates where domain IIβ (residues 410–474) associates across 2 TnsC protomers on the C-terminal face of TnsC. **c**, Side view of TnsB–TnsC interactions, as indicated in **a**. Residues 475–542 correspond to the additional structured TnsB domain (highlighted with black lines) that is observed to interact between the two TnsC protomers shown in **b**. The red asterisk indicates the beginning of helix α4. Dotted lines indicate the flexible linker (not observed in our structure). **d**, TnsB–TnsC interactions near the ATP-binding pocket. Interacting residues are shown and dashed lines represent hydrogen bonds. **e**, Alterations to TnsB domain IIβ decrease ATP hydrolysis activity. The hydrolysis rate (v_o) is shown for each variant. Data are mean ± s.d. ($n = 3$ biological triplicates). **f**, In vitro transposition assay testing TnsB mutants in the presence (dark grey) or absence (light grey) of S15. The number of transformants is plotted for each condition tested as a proxy for the transposition activity. Data are mean ± s.d. ($n = 3$ biological triplicates). Raw data points are shown in red. The on-target percentage of transposition was

estimated from Illumina sequencing and indicated on the corresponding bar plot. **g**, The ATP-binding pocket of TnsC with cryo-EM density (grey with transparent surface), magnesium ion (green sphere) and hydrogen bonds (dashed lines) shown.

TnsB domain IIβ stimulates ATPase activity

In the *Sh*CAST system, TnsB is recruited to the target site via TnsC[17] (Fig. 1a, left). This is intrinsically tied to ATP hydrolysis, because substitution with a non-hydrolyzable ATP analogue (AMP-PNP) inhibits TnsB-mediated filament disassembly and results in non-targeted transposition[18]. Therefore, a crucial role of TnsB is also to stimulate the ATPase activity of TnsC in the process of target-site selection. In prototypic Tn7, C-terminal fragments are sufficient to stimulate TnsC ATPase activity[28]. In *Sh*CAST, two TnsB–TnsC interactions have been biochemically demonstrated to be required for this process, one of which corresponds to $\text{TnsB}^{\text{hook}}$ and the other is located elsewhere in full-length TnsB[17]. Notably, within the transpososome structure, we observe a second TnsB–TnsC interaction at the C-terminal face of TnsC (Fig. 3c,d). In particular, TnsB residues 410–542 (including part of a flexible linker) are positioned to bridge two TnsC protomers (Fig. 3c). Notably, domain IIβ (residues 410–474) localizes close to the ATP-binding pocket of TnsC, abutting against helix α4 in TnsC (Fig. 3d). This helix contains functionally important, conserved residues Q185 and R189 (Fig. 3d), which recognize the nucleotide-bound state[18].

To test whether the TnsB–TnsC interactions that we observe in the transpososome structure are required for the ability of TnsB to stimulate the ATPase activity of TnsC (and trimming of the TnsC filament), we tested various mutations using a malachite green ATP hydrolysis assay (Fig. 3e and Methods). Consistent with expectations, TnsC has low basal levels of ATPase activity on its own, and wild-type TnsB substantially stimulates ATPase activity (Fig. 3e). Introduction of the Y439A mutation in TnsB reduced the ability of TnsB to stimulate ATP hydrolysis, but other mutations (such as R432A; Fig. 3d,e) had no effect. To test whether domain IIβ is generally required, we next tested various TnsB fragments for their ability to stimulate ATP hydrolysis. The C-terminal domain (CTD, residues 476–584) of TnsB or a C-terminal truncation (ΔCTD, residues 1–475) of TnsB did not

stimulate ATP hydrolysis, consistent with previous work[17]. By contrast, the C-terminal domain with the domain IIβ (IIβ+CTD, residues 410–584) increased ATP hydrolysis above basal levels of TnsC ATPase activity (that is, without TnsB) or CTD (Fig. 3e). Despite these effects on ATP hydrolysis, TnsB mutations at this interface only slightly reduced transposition activity compared with the wild type (Fig. 3f). However, consistent with the idea that these TnsB mutants have a reduced capacity to stimulate ATP hydrolysis, targeted transposition (as assessed by Illumina sequencing (Methods)) was significantly reduced in the absence of S15 (69% for the wild type versus less than 6% for the mutants) (Fig. 3f and Extended Data Fig. 6). The addition of S15, which has been shown to boost overall transposition activity[23], resulted in a less marked effect on on-site targeting (99% for the wild type versus 76–93% for the mutants) (Fig. 3f and Extended Data Fig. 6), consistent with the idea that S15 can generally promote targeted transposition by stabilizing CRISPR effector binding. It is particularly noteworthy that, in the absence of S15, wild-type TnsB retains high levels of on-site targeting (69%), whereas TnsB mutants have no targeting ability (Extended Data Fig. 6). These marked effects suggest that the ability of TnsB to promote TnsC filament disassembly is compromised, similar to the effects observed with AMP-PNP[18].

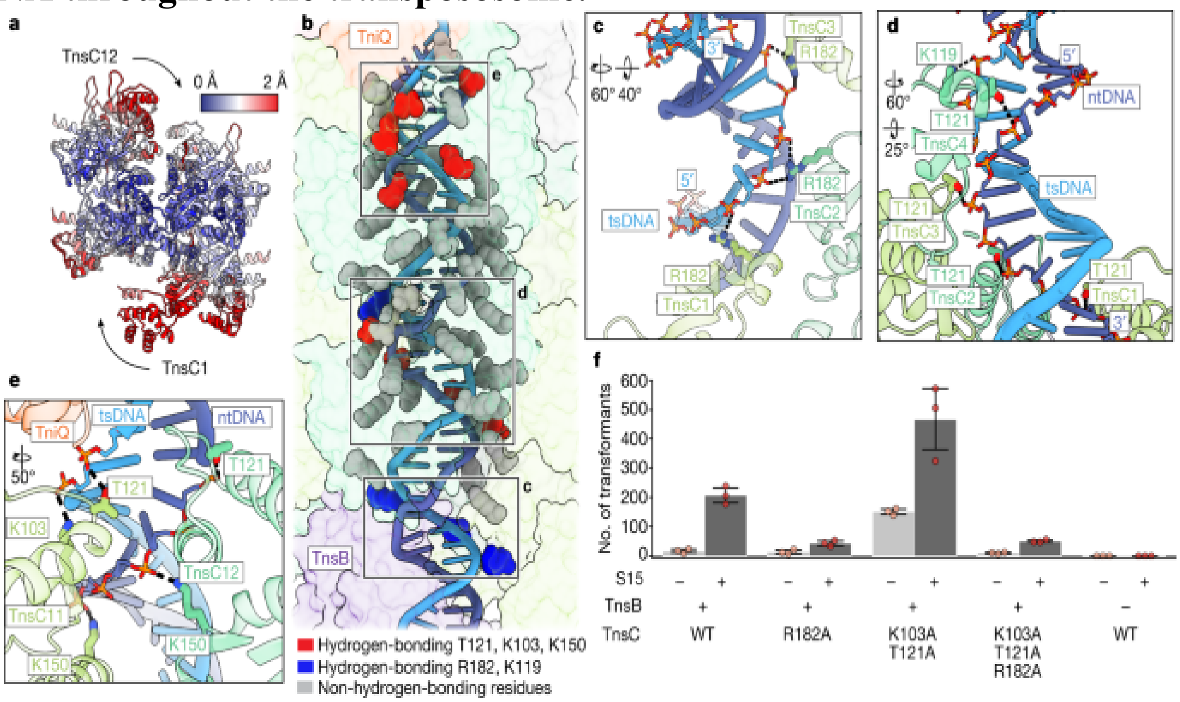
In this context, the defined assembly of TnsC that we observe in the transpososome may correspond to either an ATP hydrolysis-resistant state or a stable post-hydrolysis state. To distinguish between these two possibilities, we used local refinement (focusing on TnsC) to generate a 3.2 Å map (Supplementary Fig. 1c,d), that is sufficient to unambiguously distinguish between ATP and ADP. This improved map demonstrates that ATP is bound to all protomers (Fig. 3g and Supplementary Fig. 5). Our observations are consistent with the idea that although TnsB stimulates ATP hydrolysis and TnsC filament disassembly, aspects of the transpososome structure probably prevent hydrolysis, allowing the stable configurations observed here.

Transpososome TnsC–DNA interactions

In the absence of additional factors, ATP-bound TnsC forms continuous helical filaments (referred to throughout as helical TnsC) with DNA, and has no preference for filament length[15,18]. However, the precise insertion

profiles of *Sh*CAST suggest that TnsB trims back TnsC filaments to a specific length before transposon DNA integration[2]. Thus, we explored whether structural differences distinguish helical TnsC from the structure observed here. Although TnsC in the transpososome is globally similar to helical TnsC (Protein Data Bank (PDB) ID: 7M99; overall Cα r.m.s.d. = 1.6 Å), it is not helically symmetric (Supplementary Video 2 and Supplementary Fig. 6), because imposing helical symmetry resulted in a lower resolution reconstruction with aberrant features (Supplementary Fig. 7). The largest structural changes in the assembly correspond to the TnsB-proximal (TnsC1) and Cas12k-proximal (TnsC12) (Fig. 4a) TnsC protomers. These protomers are collapsed towards the centre of the TnsC assembly with Cα r.m.s.d. between 2.1 and 2.5 Å (Fig. 4a). This suggests that the slight conformational changes observed here are due to interactions with other CAST components within the transpososome, namely TniQ and TnsB.

Fig. 4: TnsC forms a network of important interactions with target DNA throughout the transpososome.



a, TnsC protomers are coloured by changes in Cα r.m.s.d. between the helical TnsC (PDB: 7M99) and TnsC in the transpososome. The arrows represent the inward movement of TnsC, starting from the helical filament. **b**, TnsC–DNA interactions throughout the transpososome. TnsC residues forming hydrogen-bonding interactions with DNA, at least 4 Å away, are

represented by red (K103, T121 and K150, previously identified interacting residues) or blue (R182 and K119; newly identified interacting residues) spheres. TnsC residues that are close to, but do not form specific interactions with the DNA backbone, more than 4 Å away, are represented by grey spheres. The target strand DNA (tsDNA) (light blue) and non-target strand DNA (ntDNA) (dark blue) are represented in ribbon. **c**, TnsB-proximal TnsC protomers interact with the sugar-phosphate backbone of tsDNA. **d**, TnsB-proximal TnsC protomers are shown to interact with the ntDNA backbone. **e**, TnsC protomers adjacent to Cas12k interact with both tsDNA and ntDNA immediately downstream of the R-loop. In **c**–**e**, hydrogen-bonding interactions between protein residues and the sugar-phosphate backbone of DNA are represented with dashed lines. Rotations relative to **b** are indicated. **f**, In vitro transposition assay of TnsC residues observed to interact with DNA. Data are mean ± s.d. (*n* = 3 biological triplicates). Raw data points are shown in red.

Given the lack of structural rearrangements of TnsC within the transpososome, it was notable that TnsC–DNA interactions appeared markedly different in the context of the transpososome compared with previous TnsC structures obtained in the absence of other transposition factors. In previous work, the helical symmetry of TnsC tracked with that of the bound DNA duplex[15,18]. In doing so, helical TnsC imposes its own helical symmetry, which distorts the bound DNA through underwinding[15,18]. In contrast to previous TnsC structures, within the transpososome, TnsC does not track specifically with either strand of the DNA duplex (Fig. 4b). Comparing the structure of transpososome target DNA to B-form DNA, we find that target DNA is only slightly underwound compared with B-form DNA (11 bp versus 10 bp per turn) and does not match the layer line spacing of TnsC in the transpososome (~40 Å; Extended Data Fig. 7).

Owing to the mismatch in TnsC (40 Å) and DNA (36 Å) repeat lengths, the specific TnsC–DNA interactions made in each of the two TnsC turns in the transpososome are distinct (Fig. 4b). In helical TnsC structures, residues K103 and T121 are consistently found to track with one particular strand of duplex DNA[18] (5′ to 3′ in the direction of the C-terminal to the N-terminal face, Supplementary Fig. 8), which we confirmed in improved, high-resolution (3.5 Å) cryo-EM reconstructions of TniQ–TnsC (Supplementary

Fig. 9 and 10). In the context of the transpososome, we observed a network of interactions, some of which correspond to new interactions by previously identified DNA-binding residues[15,18] (T121, K103 and K150, red) (Fig. 4b) or new interactions made by residues not previously associating with DNA (K119 and R182, blue) (Fig. 4b). The remaining interactions throughout the two turns of TnsC correspond to general electrostatic interactions, as these residues are too distant to form specific hydrogen-bonding interactions with the DNA backbone (grey residues) (Fig. 4b and Extended Data Table 2). Notably, we find that the two different sets of residues forming specific interactions with the DNA backbone are generally segregated spatially, either towards the Cas12k-proximal end (red residues) (Fig. 4b) or towards the TnsB-proximal end (blue residues) (Fig. 4b). In particular, the new interactions with the DNA backbone, involving R182 (Fig. 4c) and K119 (Fig. 4d) are supported by cryo-EM density (Supplementary Fig. 11). Residues previously shown to interact with DNA (T121 and K103) in previous structures of TnsC are now shown to associate with both target and non-target strands of the DNA duplex (Fig. 4e).

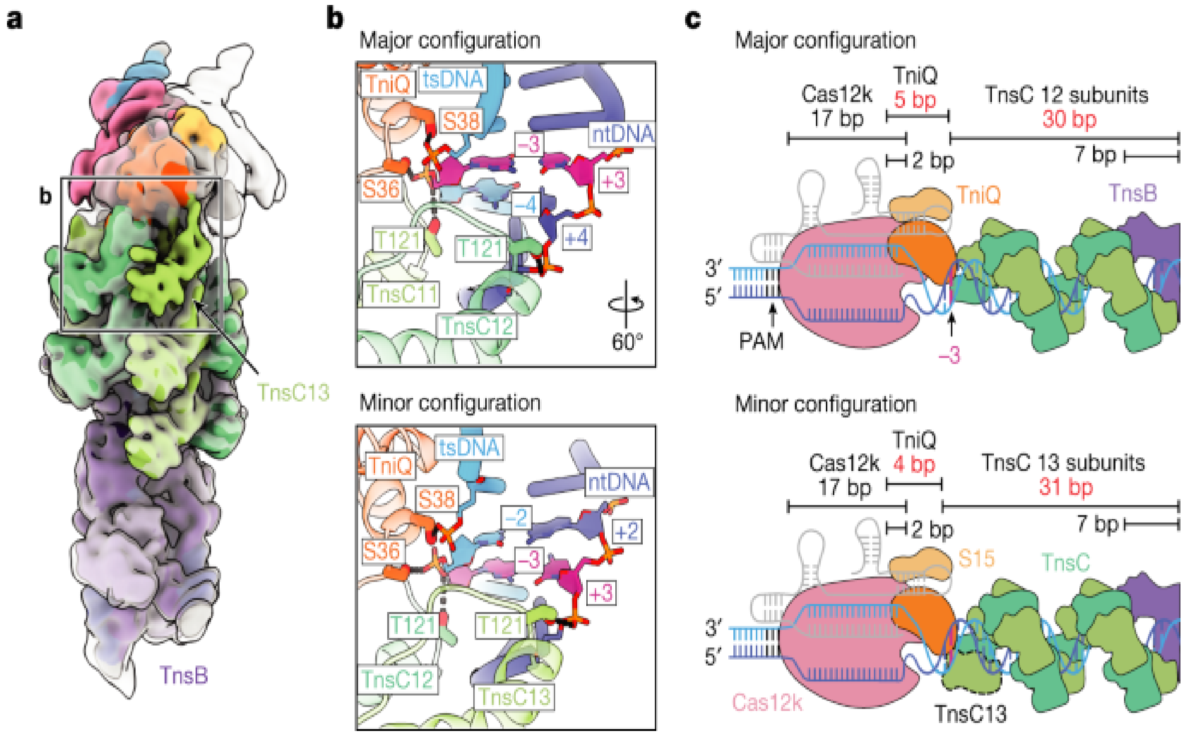
We thus hypothesized that TnsC residue R182 is critical for *Sh*CAST transposition. Consistently, both single (R182A) and triple mutants (K103A/T121A/R182A) significantly decreased the overall transposition activity (Fig. 4f), whereas the double mutant K103/T121 showed an overall increase in transposition activity (Fig. 4f), consistent with our previous results[18]. Mutant phenotypes were consistent regardless of whether S15 is added (Fig. 4f), further emphasizing the importance of R182. We hypothesize that this network of interactions corresponds to the distinct roles of TnsC residues in transpososome formation and stabilization. Furthermore, because these interactions are not observed in the TniQ–TnsC structure (Supplementary Fig. 10), this is not solely owing to the binding of TniQ, but must be related to R-loop formation, consistent with other reports[23].

The basis of insertion spacing variability

We were able to distinguish two different transpososome populations containing either 12 or 13 TnsC protomers (the former is shown in Fig. 1). Because particles containing 12 TnsC protomers were most frequently observed (73% of particles used in final refinement) and yielded a slightly

higher resolution cryo-EM map (3.5 Å versus 3.7 Å; Extended Data Table 1), we refer to this configuration as the major configuration. Comparing the two configurations, the specific TnsC–DNA interactions in protomers adjacent to TnsB are virtually identical, since the same interactions are made at the same positions on the target DNA substrate (Extended Data Fig. 8). Aligning both transpososome reconstructions (major and minor configurations) on the basis of TnsB reveals an additional TnsC protomer (TnsC13) next to Cas12k in the minor configuration, which contains an additional TnsC protomer compared with the major configuration (12 TnsC protomers) (Fig. 5a). Correspondingly, Cas12k is rotated by approximately 60° in the minor configuration (solid surface, Fig. 5a) compared with the major configuration (transparent surface, Fig. 5a). At first glance, the TnsC–DNA contacts adjacent to Cas12k also appear similar. However, these contacts are shifted over by one nucleotide in the minor configuration compared with the major configuration (Fig. 5b). That is, the last two TnsC protomers in both configurations interact in the same way with TniQ, resulting in a translational shift of TnsC–DNA interactions (Fig. 5b,c). This leads to slight variability in the number of base pairs contacted by TniQ (5 bp (major configuration) versus 4 bp (minor configuration)) (Fig. 5c) and TnsC (30 bp versus 31 bp, respectively) (Fig. 5c). The structural variations described here suggest that more than one transpososome complex is stable, leading to an elegant mechanism to explain the slight variability (around 5 bp) in *Sh*CAST insertion profiles[2].

Fig. 5: Two different transpososome configurations differ at the Cas12k-proximal end.



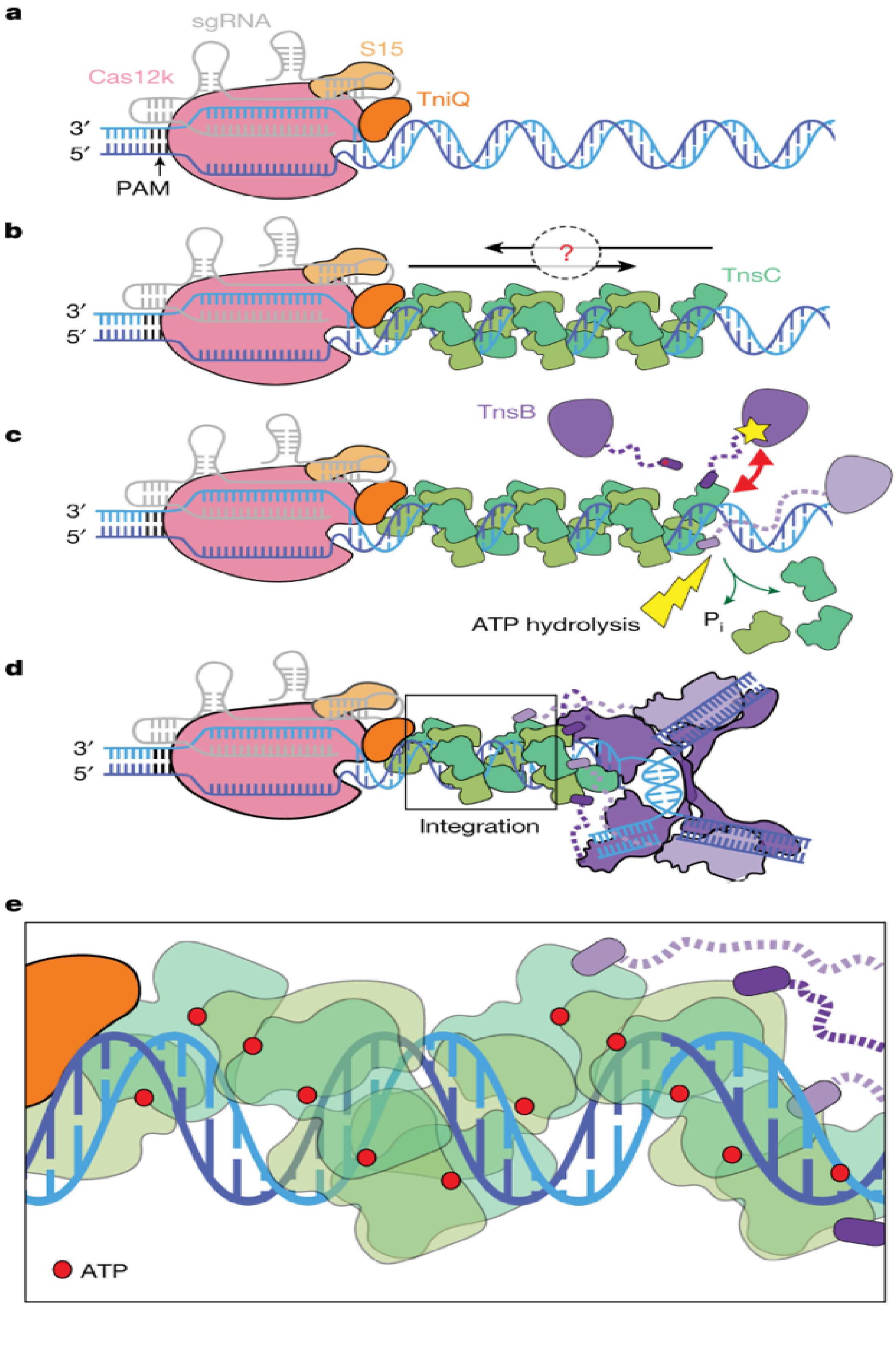
a, The cryo-EM reconstructions (low pass filtered to 10Å) of major and minor configurations of transpososome. The reconstructions were aligned with respect to TnsB, where protein–DNA interactions are identical in the two configurations. The map corresponding to the minor configuration is coloured and opaque, and the major configuration is transparent. The outline indicates TnsC–DNA interactions at the Cas12k face. **b**, Expanded view of the outlined region in **a**, showing that major and minor configurations interact with both target-strand DNA (tsDNA) and non-target strand DNA (ntDNA). Residues from TniQ and TnsC in the minor configuration interact with the phosphate backbone that are shifter by 1 bp towards the R-loop compared with the major configuration. Hydrogen-bonding interactions between protein residues and the DNA backbone are represented with dashed lines. Base pairs interacting with the final two protomers of TnsC are represented in stick. Positions three bases downstream of the R-loop (−3 on tsDNA and +3 on ntDNA) are coloured magenta as a landmark, facilitating comparison of the two configurations. **c**, Schematic of DNA-binding contacts of each component (Cas12k, TniQ and TnsC) in major (top) and minor (bottom) configurations. The number of base pairs contacted by each component is shown and the length is indicated with bar-ended lines. Positions interacting with two protein components (Cas12k and TniQ or TnsC and TnsB) are indicated as overlapping bars. The minor configuration

includes and additional TnsC protomer (TnsC13, dashed outline) proximal to Cas12k, which makes the DNA-binding contribution of TniQ and TnsC different in the two configurations (highlighted in red). The third base pair downstream of the R-loop is coloured magenta, as in **b**.

Discussion

Our structural observations enable us to fill in crucial mechanistic details of the recruitment and insertion of transposon DNA at a CRISPR-defined target site (Fig. 6). Once a productive complex is formed consisting of target-site proteins Cas12k, S15, TniQ and TnsC (Fig. 6a,b), TnsB–TnsC interactions stimulate ATP hydrolysis and TnsC filament disassembly. Our structure also suggests that the TnsB–TnsC contacts required for filament disassembly involve two TnsB domains, $TnsB^{hook}$ and domain IIβ (Fig. 6c). ATP hydrolysis (and TnsC filament disassembly) continues until TnsB encounters the two TnsC turns closest to Cas12k. We hypothesize that two turns of TnsC are required for transpososome formation with the *Sh*CAST element (Fig. 6d), a state resistant to TnsB-mediated ATP hydrolysis (Fig. 6e) and stabilized by new TnsC–DNA contacts, among them K103, T121, K119 and R182. These contacts appear to stabilize a TnsC configuration that is not fully engaged with DNA compared to the helical TnsC formed outside the transpososome. We speculate that the two turns of TnsC that resist TnsB-promoted disassembly might be a basis for the specific activation of TnsB at the target site. However, the detailed mechanism whereby TnsC activates ATPase activity, resists filament disassembly and stimulates integration requires further investigation.

Fig. 6: Mechanistic model of recruitment and integration of the transposition components in *Sh*CAST.


a
sgRNA
S15
Cas12k
TniQ
3′
5′
PAM
b
?
TnsC
3′
5′
c
TnsB
3′
5′
ATP hydrolysis
Pi
d
3′
5′
Integration
e
ATP

a, The CRISPR effector Cas12k (pink) bound to sgRNA (grey) defines the target site by forming an R-loop and associating with S15 (tan) and TniQ (orange). **b**, TnsC (green) may polymerize on DNA (blue) in two directions: towards or away from the target site (indicated by the black arrows and the question mark). TnsC interacts with target-site-associated proteins (Cas12k–sgRNA, S15 and TniQ) to form the recruitment complex. **c**, Two interactions between TnsB (purple) and TnsC: TnsB is recruited to TnsC filaments by $TnsB^{hook}$ (purple, indicated with a red asterisk) and domain IIβ (indicated approximately by a yellow star) from TnsB can interact with TnsC (indicated by a red double arrow), lead to disassembly of TnsC and promote ATP hydrolysis (yellow lightning bolt), resulting in the release of phosphate (green arrows). Dashed lines (purple) represent flexible linkers between $TnsB^{hook}$ and the rest of TnsB. **d**, TnsB is recruited to the target site and forms the STC upon integration. Four $TnsB^{hook}$ are bound to four TnsC protomers proximal to TnsB through the flexible linker (dashed purple line). Two turns of TnsC (12 protomers) are stably formed against disassembly. Together with target-site-associated proteins and TnsB STC, all CAST components form the transpososome at the integration site. **e**, A magnified view of TnsC (boxed) in the transpososome shown in **d**. All TnsC protomers are in the ATP-bound state (ATP is shown as a red circle). TnsC in the transpososome does not specifically track with DNA helical symmetry, unlike previous helical structures of TnsC.

This structure also reveals how *Sh*CAST transpososome architecture conserves TnsC function across Tn7-like elements. *Sh*CAST TnsC forms continuous helical filaments on DNA[15,18], which are distinct from TnsC configurations found in the prototypic Tn7[19] and I-F3 CAST subfamily[11]. TnsC from prototypic Tn7 forms heptameric oligomers[19] and led to the original hypothesis that the ring would allow segregated interaction interfaces: the TnsD interface maps to the N-terminal face, and the TnsA–TnsB interface maps to the C-terminal face[19]. Although the I-F3a CAST TnsC adopts a heptameric structure[11], it remains unknown whether TnsC in this system has similarly segregated interaction interfaces. However, chromatin immunoprecipitation–sequencing experiments suggest that TnsC has a key role in associating with the Cascade complex, serving as a selectivity factor in subsequent recruitment of TnsA–TnsB[11]. Therefore, despite the structural differences in *Sh*CAST TnsC, we hypothesize that the

segregated interactions that we observe in the *Sh*CAST transpososome are conserved across Tn7-like transposition systems.

The segregation of interactions that we observe in the transpososome also suggests a mechanism for how the core transposition machinery can co-opt different CRISPR effectors via TniQ. This is particularly interesting, considering the I-B CAST elements, which contain features of both prototypic Tn7 and CAST elements[6]. TnsC and the cognate transposase in I-B CAST elements can associate with two different TniQ domain proteins: TnsD (sequence-specific DNA-binding protein) or TniQ (associated with a CRISPR effector) to direct insertions[6]. Competition between these two targeting modalities suggests that a single interface on TnsC can interact with either TniQ or TnsD[6]. Our structure provides an elegant explanation for how this is achieved: by separating targeting machinery (TniQ or TnsD) from the core transposition machinery (TnsA–TnsB) via segregated interfaces on TnsC. This ensures that co-option of target-site-binding proteins (that is, with different CRISPR effectors or different DNA-binding proteins) does not interfere with the function of core transposition machinery.

Although the transpososome structure is remarkably uniform and well-resolved (Extended Data Fig. 2) compared with the heterogeneity of other reconstitutions[23], we observe a slight variability in the oligomeric composition of TnsC within transpososome populations. This is not altogether unexpected, since CAST insertion profiles also contain a slight variability (around 5 bp) in their insertion profiles[2]. Notably, the protein–protein associations that we observe in either configuration are essentially identical, except that TnsC and TniQ are shifted one nucleotide closer to Cas12k (Fig. 5c). Therefore, transpososome associations are constructed to allow minor flexibility. Taking this a step further, it is possible that the basis of CAST insertion variability lies in the stable configurations of TnsC that preserve CAST component interactions. This suggests that more precise CAST elements can be engineered by reinforcing (that is, stabilizing) specific spatial associations between CAST components within the transpososome.

Methods

Protein cloning of *Sh*CAST mutants and S15

TnsC (K103A, T121A, and R182A), TnsB (ΔCTD; residues 1–475, CTD; residues 476–584, IIβ+CTD; residues 410–584, R432A, and Y439A), and TniQ (Δ1–10; residues 11–167, W10A, F12A, and H34A) mutants were cloned from pXT130-TwinStrep-SUMO-ShTnsC (Addgene #135526), pXT129-TwinStrep-SUMO-ShTnsB (Addgene #135525) and pXT131-TwinStrep-SUMO-ShTniQ (Addgene #135527), respectively, using the Q5 site-directed mutagenesis kit (NEB) and two corresponding primers (forward and reverse primer pair are listed in Supplementary Table 1). Double or triple mutants were generated by introducing mutation on top of the mutant plasmid vectors. For S15, the DNA sequence for *Escherichia coli* S15 was amplified by PCR from *E. coli* K-12 MG1655 chromosomal DNA using the following primers: RPS15_Ext_Hifi_F and RPS15_Ext_Hifi_R (Supplementary Table 1). This gene fragment was cloned into a linearized vector backbone (original ShTniQ vector from Addgene #135527) containing TwinStrep-SUMO at the N terminus of S15 using the NEBuilder HiFi assembly protocol (NEB). TnsC clones were transformed into *E. coli* BL21(DE3) competent cells (NEB), while TnsB and S15 clones were transformed into homemade T1 phage-resistant pRIPL BL21(DE3), derived from transforming pRIPL into BL21(DE3) Competent *E. coli* (NEB).

Protein expression and purification of *Sh*CAST components and S15

All clones except S15 were purified using previously described protocols[2,18]. For purification of S15, single colonies were inoculated in 10 ml LB medium containing 34 μg ml^{-1} chloramphenicol and 100 μg ml^{-1} ampicillin, and grown overnight at 37 °C as starter cultures. The overnight starter culture was added to 1.5 L of TB medium containing the same ratio of antibiotics and grown until an absorbance of 0.4 at 37 °C. The temperature was lowered to 20 °C and the culture was induced at an absorbance of 0.6 with 0.4 mM isopropyl-β-d-thiogalactopyranoside and continued to grow overnight (16–18 h). Cells were collected by centrifugation at 5,000 rpm for 10 min and resuspended in lysis buffer (50 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol, 1 mM DTT, 1 mM PMSF, 0.1% NP-40 and protease inhibitors (Pierce, ThermoScientific). The cell

lysate was sonicated for 15 min with a pulse on for 2 s with 10 s rest in between. The sonicated lysate was centrifuged for 30 min at 10,000 rpm at 4 °C. The supernatant was loaded onto a Strep-Tactin Superflow resin column (Qiagen) equilibrated with loading buffer (50 mM Tris pH 7.4, 500 mM NaCl, 1 mM DTT, and 5% glycerol). The column was washed with 5 column volumes (~15 ml) of wash buffer (50 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol) before eluting with elution buffer (50 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol and 4 mM desthiobiotin (SigmaAldrich)). The consolidated eluate was then incubated overnight at 4 °C with SUMO protease (1/100 mass ratio of SUMO protease to S15 protein).

S15 was next purified using a heparin purification step. Protein from the Strep-Tactin elution (in 500mM NaCl) was diluted to a final salt concentration of 200 mM NaCl before loading onto a 1 ml heparin column (GE Healthcare) equilibrated in low-salt buffer (50 mM Tris pH 7.4, 200 mM NaCl, 1 mM DTT and 5% glycerol). The heparin column was washed with 20 column volumes of low-salt buffer, followed by a gradient elution from 200mM NaCl to 1 M NaCl using high-salt buffer (50 mM Tris pH 7.4, 1.2M NaCl, 1 mM DTT, and 5% glycerol). Fractions containing S15 were concentrated, and buffer exchanged before size-exclusion chromatography was performed using a Superdex 200 increase 10/300 (Cytiva) equilibrated in the following buffer: 20 mM HEPES pH 7.5, 250 mM KCl, 1 mM DTT, and 5% glycerol. Fractions containing S15 were consolidated and concentrated to ~2 mg ml^{-1} using a 6 ml 3K molecular weight cut-off centrifugal concentrator (ThermoFisher), before being aliquoted and flash-frozen in liquid nitrogen for long-term storage at −80 °C.

Preparation of DNA substrate

The general idea of the substrate design for the STC was described previously[17]. However, instead of a symmetric strand-transfer DNA, we designed an asymmetric DNA that mimics the Cas12k-guided transposition product (Supplementary Table 1). This design includes two DNA substrates: (1) target-pot DNA containing left-end (LE) sequence and protospacer and (2) donor pot DNA containing right-end (RE) sequence. Each DNA substrate was prepared independently. First, target-pot DNA substrate was prepared by annealing four synthetic oligonucleotides (IDT) of the

following: target-LE_F, non-target_R, LE_R, and 5′ dethiobiotinylated LUEGO[29] (Supplementary Table 1). These four oligonucleotides were mixed in a molar ratio of 10:11:11:10. Second, donor pot DNA is composed of three synthetic oligonucleotides including: RE_F, RE_R1 and RE_R2 (Supplementary Table 1). These oligonucleotides were mixed in a 10:11:11 molar ratio. The mixture of oligonucleotides was supplemented with a 10× concentrated annealing buffer to make the final buffer composition the following: 10 mM Tris pH 7.5, 50 mM NaCl, and 1 mM EDTA. The mixture was then heated up to 95 °C for 5 min and cooled down to 30 °C at the rate of 1 °C per minute using a thermal cycler (BioRad).

Preparation of sgRNA

Using two primers, sgRNA_For and sgRNA_PSP1_Rev (Supplementary Table 1), the DNA sequence encoding the T7 RNA polymerase promoter and sgRNA was amplified by PCR from pHelper_ShCAST_sgRNA vector (Addgene, #127921). The DNA template was then subjected to GeneJet PCR purification (ThermoScientific). sgRNA was produced by in vitro transcription using the HiScribe T7 High Yield RNA synthesis kit (NEB) with the PCR amplified DNA sequence as the template. Purified sgRNA was aliquoted and stored at −20 °C.

Reconstitution of transpososome complex and recruitment complex

The transpososome complex and recruitment complex were generated by stepwise assembly using a pull-down assay on streptavidin beads (Fig. 1b). For the target pot, annealed target-pot DNA was first bound to Streptavidin Mag Sepharose magnetic beads (Cytiva) equilibrated with reaction buffer (20 mM HEPES pH 7.5, 250 mM KCl, 15 mM $MgCl_2$, and 1 mM DTT) and incubated at room temperature with shaking for 30 min. Cas12k and sgRNA were reconstituted in a 1:2 molar ratio and incubated for 30 min at 37 °C in the following buffer: 20 mM HEPES pH 7.5, 160 mM NaCl, 15 mM $MgCl_2$, and 1 mM DTT. The Cas12k–sgRNA complex was added to the target-pot DNA-bound beads and incubated for 30 min at 37 °C in the presence of a fivefold molar excess of S15 to DNA. The beads were washed three times with 500 µl wash buffer (20 mM HEPES pH 7.5, 250 mM KCl, 15 mM

$MgCl_2$, 0.05% Tween20, and 1 mM DTT) to remove excess protein and nucleic acids. Twentyfold molar excess of TniQ and tenfold molar excess of S15 to DNA were added to the beads and incubated at 37 °C for 30 min. Tenfold molar excess of TnsC was diluted to a final salt concentration of 250 mM NaCl and added to the TniQ–S15 beads solution containing a final concentration of 1 mM ATP. The sample was incubated for 30 min at 37 °C before the beads were washed three times with 500 µl wash buffer containing 1 mM ATP. The recruitment complex was prepared by eluting the target-pot DNA at this point using 50 µl of elution buffer (20 mM HEPES, 250 mM KCl, 15 mM $MgCl_2$, 0.05% Tween20, 10 mM biotin, 1 mM ATP, and 1 mM DTT). For transpososome assembly, donor pot containing TnsB and donor pot DNA was prepared independently in 6:1 molar ratio with the following final buffer composition: 25 mM HEPES pH 7.5, 100 mM NaCl, 15 mM $MgCl_2$, and 1 mM DTT; and incubated at 37 °C for 30 min. The reconstituted donor pot was added to the beads and incubated for 40 min at 37 °C. Three washes of 500 µl wash buffer containing 1 mM ATP were performed before eluting the transpososome complex from the beads with 50 µl elution buffer. The eluate was diluted with the wash buffer containing 1 mM ATP for cryo-EM or negative-staining EM sample preparation by 6–10 fold or 15 fold respectively.

In vitro transposition assay and high-throughput mapping of transposition events

All purified proteins (except Cas12k) were diluted to 2.5 µM using stock buffer (500 mM NaCl, 50 mM Tris pH 7.4, 10% glycerol, 0.5 mM EDTA, 1 mM DTT). Cas12k was assembled with sgRNA before incubation with other components in the following manner: sgRNA was added to Cas12k to make 2.5 µM Cas12k and 30 µM sgRNA. Separate reactions were prepared for the target pot and donor pot. The target pot contains 2.24 nM pTarget_pBC_KS+_PSP1 (a gift from the J. E. Peters laboratory), 104 nM of Cas12k, 104 nM TnsC, 104 nM TniQ and 1.25 µM sgRNA in transposition buffer (26 mM HEPES, 50 mM KCl, 0.2 mM $MgCl_2$) supplemented with 2 mM DTT, 50 $\mu g\ ml^{-1}$ BSA and 2 mM ATP. The donor pot contains 1.08 nM pDonor_ShCAST_kanR (Addgene #127924), 104 nM TnsB, 2 mM DTT, 50 $\mu g\ ml^{-1}$ BSA and 2 mM ATP in transposition buffer. Both pots were

incubated at 37 °C for 30 min before combining and adding $MgOAc_2$ to a final concentration of 15 mM. The combined reaction was further incubated at 37 °C for 2 h. After the final incubation step, 20 µl of the reaction mixture was digested by 1 µl of Proteinase K (Ambion) and incubated at 37 °C for 1 h. 10 µl of the reaction was then transformed into 100 µl of Stellar competent cells (Takara Bio) and plated on kanamycin plates. All the experiments were done in biological triplicates.

From the plates of each condition, the colonies were washed up and diluted to absorbance of 0.5 using LB. The liquid culture was then grown for 3 h at 37 °C before the plasmid extraction using a miniprep kit (QIAGEN). Prepared DNA was sequenced by the Microbial Genome Sequencing Center (https://www.seqcenter.com/) using the Illumina DNA sequencing service on the NextSeq 2000 platform. Paired-end reads (2 × 151 bp) were analysed using BBtools v38.98 (http://sourceforge.net/projects/bbmap/). First, the reads were processed by BBDuk (BBtools v38.98) to collect adjacent target DNA sequences from all the reads containing 30bp of the *Sh*CAST left-end sequence. Then these sequences were mapped onto the pTarget_pBC_KS+_PSP1 using BBMap (BBtools v38.98) with a 90% sequence identity cut–off. Number of base pairs between the end of the PAM and the beginning of the *Sh*CAST left-end sequence was used to define the position of the transposition. Based on the mapping, the transpositions at 55–70 bp downstream of the PAM were considered to be on-target transposition. The number of on-target transposition reads was divided by the total number of transposition reads to estimate the on-target ratio.

Malachite green ATP hydrolysis assay

One-hundred millilitres of malachite green reagent was prepared by the following protocol. First, 34 mg of malachite green carbinol base (ChemCruz) was dissolved into 40 ml of 1N HCl. 1 g of Ammonium Molybdate (MacronChemicals) was dissolved in a separate 14 ml of 1N HCl. These two solutions were mixed and diluted up to 100 ml using ddH_2O. The reagent was then filtered using 0.45 µm syringe filter (VWR), and covered with an aluminum foil to avoid light. ATPase assay was completed using purified protein stocks of TnsC, and constructs of TnsB (wild type, ΔCTD, CTD, IIβ+CTD, R432A and Y439A). To minimize the

processing time, 2× protein pot and 2× ATP pot were prepared independently. Buffer composition for both pots was the same as follows: 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT. In addition to the described buffer, 2× protein pot contains 10 µM TnsC, 2 µM TnsB, and 2.5 µM 60 bp DNA (annealed using 60bp_top and 60bp_bot, Supplementary Table 1), and 2× ATP pot contains 1 mM ATP and 2 mM $MgCl_2$. One-hundred microlitres of each pot was mixed to initiate the ATP hydrolysis reaction. For each time point (10, 20, 30, 40, 50, 60, 90 and 120 min), 20 µl of the reaction mix was transferred to the 96-well plate (Corning) that contains 5 µl of 0.5 M EDTA to quench further hydrolysis. After taking all the time points, 150 µl of the room-temperature Malachite Green reagent was added to each well, and the plate was shaken for ~5 min to develop the colour for imaging. Absorbance at 650 nm was read from each well using the i-control v1.101.4 software on an Infinite 200 plate reader (TECAN). For calibration, KH_2PO_4 solutions of the following concentrations were used to generate a calibration curve: 0 µM, 4 µM, 8 µM, 12 µM, 16 µM, 24 µM, 40 µM and 60 µM. The data was plotted with time as the x-axis and the concentration of released inorganic phosphate as the y-axis. The slope was obtained from the linear regression to get the reported v_o values (in moleculas of ATP per min). Three independent experiments were done for each condition, to generate the reported bar plot.

Cryo-EM sample preparation and freezing of transpososome complex

The transpososome complex sample was prepared by diluting the elution from the pull-down by 6-, 8- and 10-fold using wash buffer (described above) containing 1 mM ATP. Graphene oxide-coated grids were prepared following the protocol described previously[17,30,31]. Four microlitres of transpososome complex sample was loaded on the carbon side of a graphene oxide-coated grid and incubated for 20 s in the Mark IV vitrobot chamber (ThermoFisher), which was set to 4 °C and 100% humidity. Each grid was blotted for 6 s with a blot force of 5, and then plunged into liquid ethane cooled with liquid nitrogen.

Cryo-EM imaging and image processing of transpososome complex

Vitrified samples of transpososome complex were first screened using Talos Arctica (ThermoFisher) operated at 200 kV prior to large-scale data collection. Screened grids were imaged using a Titan Krios (300 kV, ThermoFisher) equipped with a BioQuantum energy filter (Gatan) and K3 direct electron detector (Gatan). A total of 15,740 micrographs were collected using Leginon v3.5[32] at 81,000× magnification (1.067 Å per pixel) using image shift, with the nominal defocus from −0.8 µm to −2.5 µm. Each movie was collected with 2 s of exposure with a total dose of 49.91 electrons per Å^2, fractionated into 50 frames. Frames were aligned using MotionCor2[33] through Appion v3.4[34], which was then imported to cryoSPARC v3.3.1[35] for contrast transfer function (CTF) estimation and downstream image analysis. The workflow described below is shown in Extended Data Fig. 2. Using a template picker from cryoSPARC v3.3.1, initial particle stack was subjected to 2D classification in cryoSPARC v3.3.1. Two-dimensional classification resulted in 536,450 particles from selected 2D averages, which was subjected to ab initio reconstruction in cryoSPARC v3.3.1. Ab initio reconstruction separated particles into two classes, one containing 53% (285,090 particles) and another containing 47% (250,842 particles). The two classes were then subjected to homogenous refinement in cryoSPARC respectively. The particle stacks from both classes were subjected to 3D classification in RELION v4[36,37], which removed junk particles with weak densities of Cas12k or TnsB. Each classification resulted in a particle stack that had well-defined configurations of *Sh*CAST transpososome. The final particle stacks (188,055 particles for major configuration of transpososome and 67,096 particles for minor configuration of transpososome) were subjected to iterative CTF refinements[38] and Bayesian polishing[39] to improve resolution. The final maps of both major and minor configurations were then subjected to local refinement in cryoSPARC v3.3.1. Volume maps of Cas12k–S15-TniQ only, TnsC only, and TnsB only were generated in UCSF Chimera v1.14[40] using the volume eraser tool to remove the map. A mask for Cas12k–S15–TniQ, TnsC and TnsB, respectively, was then generated in RELION v4[36] using the volume produced by Chimera as the input. Local refinement was then done in

cryoSPARC v3.3.1[35] with the mask generated by RELION v4 and the full map without particle subtraction. Local refinement of the major configuration of the transpososome resulted in 3.1 Å for the Cas12k–S15–TniQ region, 3.2 Å for both the TnsC and TnsB region. Local refinement of the minor configuration of the transpososome resulted in 3.6 Å for the Cas12k–S15–TniQ region, 3.6 Å for TnsC, and 3.8 Å for the TnsB region.

For Fig. 2, the composite map was generated using UCSF Chimera v1.14[40] command 'vop maximum' with the aligned reconstructions from local refinements of Cas12k and TnsC. 3D Variability in cryoSPARC v3.3.1[41] was performed on the major configuration of the transpososome using the mask and particles from the final refinement. The filter resolution was set to 6 Å and the number of modes to solve was set to 3. After the job was completed, the 3D variability display job using the particles and volumes from the 3D Variability job was completed with the output mode set to simple with 20 frames. The first series was then visualized in UCSF Chimera v1.14[40] using the "vop morph" command with all 20 frames. The results are presented in Supplementary Video 1.

To impose helical symmetry on the two turns of TnsC in the transpososome, a mask was first created using a map of TnsC only (generated by the same protocol described above) as input in cryoSPARC v3.3.1[35]. Helical refinement was then done in cryoSPARC v3.3.1[35] with the mask and the full map (major configuration) without particle subtraction. Initial helical parameters of the ATPγS-bound TnsC filaments (PDB: 7M99, rise = 6.82 Å and twist = 60º)[18] were used. Helical parameters were refined with the following range: 6.14 Å to 7.50 Å for the helical rise, and from 57º to 63º for the twist.

Model building of transpososome complex

Initial models from previous studies, including TnsC (PDB: 7M99)[18], TnsB (PDB: 7SVW)[17] and sgRNA (PDB: 7PLA)[15], were first docked into the cryo-EM density and manually rebuilt using Coot v0.9.8.2[42]. Additional models were created using AlphaFold2[43] for the following components: Cas12k, TniQ and S15, and rigid-body docked into the density. Coot v0.9.8.2 was used to manually remodel or rebuild sections of the model.

Real space refinement of the DNA substrate was completed in Phenixv1.19.1-4122[44] with both base-pair and secondary structure restraints enforced. For measuring the interface area between TniQ and TnsC protomers, UCSF ChimeraXv1.2.5[45] command 'measure buriedarea' was used with the desired two chains as inputs.

Model validation of transpososome complex

Map-model Fourier shell correlation (FSC) was computed using Mtriage in Phenixv1.19.1-4122[44]. Map-model FSC resolution of each dataset was estimated from Mtriage FSC curve, using 0.5 cut-off. The masked cross-correlation (CCmask) from Mtriage was reported as representative model-map cross-correlation. Model geometry was validated using MolProbity v4.5.1[46]. All the deposited models were submitted to MolProbity v4.5.1 server to check the clashes between atoms, Ramachandran plot, bond angles, bond lengths, side-chain rotamers, CaBLAM and C-beta outliers. All the model validation stats are summarized in Extended Data Table 1.

Helical parameter estimation using Rosetta

Helical parameters are estimated between the adjacent TnsC protomers within the complexes of the following: ATPγS-bound TnsC helical filament (PDB: 7M99), major configuration, and minor configuration of the transpososome. For example, from the major configuration of transpososome, helical parameters were estimated in between TnsC1 and TnsC2, between TnsC2 and TnsC3, between TnsC3 and TnsC4, and so on. Rosetta tool 'make_symmdef_file.pl' was used for each pair of the TnsC protomers, with an example command presented at the bottom of this section. Obtained helical rises and turns were averaged to plot a bar-graph presented in Supplementary Fig. 6.

The Rosetta command that we used for helical parameter is Rosetta/main/source/src/apps/public/symmetry/make_symmdef_file.pl -m HELIX-p input.pdb-a A-b B.

Reconstitution of $TnsB^{CTD}$–TnsC–TniQ complex and imaging for cryo-EM

The C-terminal 109 residues from wild-type TnsB (termed hereafter as $TnsB^{CTD}$) was cloned from pXT129_TwinStrep-SUMO-ShTnsB vector (Addgene, #135525) using Q5 site-directed mutagenesis kit (NEB) and two primers of following: TnsB_CTD_For and NdeI_primer_Rev (Supplementary Table 1). Cloned $TnsB^{CTD}$ was transformed to *E. coli* BL21-RIPL competent cells (Agilent) and purified following the protocol for TniQ purification. DNA substrate was prepared by annealing two synthetic oligonucleotides of BCQ_top and BCQ_bot (Supplementary Table 1). TniQ and TnsC were buffer exchanged into 25 mM HEPES pH 7.5, 200 mM NaCl, 2% glycerol, and 1 mM DTT (reaction buffer) prior to reconstitution of the complex. TnsC filament was first reconstituted by supplementing the reaction buffer with the following components: 128 µM TnsC, 8 µM DNA, 2 mM ATP, and 2 mM $MgCl_2$. After 5 min of incubation on ice, sixfold molar excess of TniQ was added to TnsC filaments and incubated at 37 °C for 1 h. $TnsB^{CTD}$ was then added to the TniQ–TnsC mixture (4:1 molar ratio of $TnsB^{CTD}$ to TnsC), followed by incubation at 37 °C for 40 min. Four microlitres of the $TnsB^{CTD}$–TnsC–TniQ sample was loaded on UltrAuFoil R1.2/1.3 gold grids (Quantifoil) and vitrified using the Mark IV Vitrobot (ThermoFisher) set to 100% humidity and 4 °C. Samples were blotted for 7 s with blot force 5, and then plunged into liquid ethane cooled with liquid nitrogen. Vitrified grids were imaged using Talos Arctica (ThermoFisher, 200 kV) with K3 direct detector (Gatan) and BioQuantum energy filter (Gatan), at 63,000× nominal magnification (1.33Å per pixel). SerialEM v4.0[47] was used for data acquisition with 3 × 3 image shift, and nominal defocus range from −1 µm to −2.5 µm. Total dose was set to 50 e^{-} $Å^{-2}$, which was fractionated into 50 frames.

Image processing and model building for $TnsB^{CTD}$–TnsC–TniQ complex

Beam-induced motion correction, CTF estimation, and initial particle picking were done using Warp v1.0.9[48] for the collected 1,271 movies.

Initial particle stack from Warp was subjected to 2D classification cryoSPARC v3.3.1[35], to get subset of TniQ-bound particles to train topaz neural network[49], which resulted in the 795,631 particles. 2D classification resulted in 624,597 particles from selected 2D averages with high-resolution features, which were subjected to heterogeneous refinement in cryoSPARC v3.3.1[35]. One class with better resolved TniQ (34%, 214,291 particles) was selected for downstream non-uniform refinement in cryoSPARC v3.3.1[35]. The particle stack was subjected to 3D classification in RELION v4[36,37] resulting in the intermediate particle stack with a stronger density of TniQ (70%, 150,358 particles). To improve the resolution of TniQ, this particle stack was subjected to two rounds of focused 3D classification (skipping alignments, tau fudge factor of 16), iterative CTF refinements[38] and bayesian polishing[39]. This resulted in the final dataset of 61,515 particles, with the gold standard FSC resolution of 3.5 Å. Due to the local variation of the map quality, LocSpiral[50] was used to post-process the map through COSMIC^2 [51]. For model building of TnsC and TniQ, previously published structure of TnsC (PDB: 7M99)[18], and AlphaFold2[43] generated TniQ were manually docked into the density, followed by manual editing using Coot v0.9.8.2[42] and relaxation using Rosetta relax[52].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic models are available through the PDB with accession codes: 8EA3 (major configuration), 8EA4 (minor configuration) and 7SVU ($TnsB^{CTD}$–TnsC–TniQ complex). All cryo-EM reconstructions are available through the EMDB with accession codes: EMD-27971 (major configuration), EMD-27972 (minor configuration) and EMD-25453 ($TnsB^{CTD}$–TnsC–TniQ complex).

References

1. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR–Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).

2. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).

3. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).

4. Faure, G. et al. CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).

5. Petassi, M. T., Hsieh, S. C. & Peters, J. E. Guide RNA categorization enables target site choice in Tn7–CRISPR–Cas transposons. *Cell* **183**, 1757–1771 e1718 (2020).

6. Saito, M. et al. Dual modes of CRISPR-associated transposon homing. *Cell* **184**, 2441–2453.e18 (2021).

7. Benler, S. et al. Cargo genes of Tn7-Like transposons comprise an enormous diversity of defense systems, mobile genetic elements, and antibiotic resistance genes. *mBio* **12**, e0293821 (2021).

8. Bainton, R. J., Kubo, K. M., Feng, J. N. & Craig, N. L. Tn7 transposition: target DNA recognition is mediated by multiple *Tn7*-encoded proteins in a purified in vitro system. *Cell* **72**, 931–943 (1993).

9. Peters, J. E. Targeted transposition with Tn7 elements: safe sites, mobile plasmids, CRISPR/Cas and beyond. *Mol. Microbiol.* **112**, 1635–1644 (2019).

10. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernández, I. S. Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system. *Nature* **577**, 271–274 (2020).

11. Hoffmann, F. T. et al. Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. *Nature* **609**, 384–393 (2022).

12. Choi, K. Y., Spencer, J. M. & Craig, N. L. The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. *Proc. Natl Acad. Sci. USA* **111**, E2858–E2865 (2014).

13. Waddell, C. S. & Craig, N. L. Tn7 transposition: two transposition pathways directed by five *Tn7*-encoded genes. *Genes Dev.* **2**, 137–149 (1988).

14. Klompe, S. E. et al. Evolutionary and mechanistic diversity of Type I-F CRISPR-associated transposons. *Mol. Cell* **82**, 616–628.e615 (2022).

15. Querques, I., Schmitz, M., Oberli, S., Chanez, C. & Jinek, M. Target site selection and remodelling by type V CRISPR–transposon systems. *Nature* **599**, 497–502 (2021).

16. Xiao, R. et al. Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Mol. Cell* **81**, 4457–4466.e5 (2021).

17. Park, J. U., Tsai, A. W., Chen, T. H., Peters, J. E. & Kellogg, E. H. Mechanistic details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM. *Proc. Natl Acad. Sci. USA* **119**, e2202590119 (2022).

18. Park, J. U. et al. Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021).

19. Shen, Y. et al. Structural basis for DNA targeting by the Tn7 transposon. *Nat. Struct. Mol. Biol.* **29**, 143–151 (2022).

20. Yin, Z., Lapkouski, M., Yang, W. & Craigie, R. Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. *Protein Sci.* **21**, 1849–1857 (2012).

21. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).

22. Tenjo-Castano, F. et al. Structure of the TnsB transposase–DNA complex of type V-K CRISPR-associated transposon. *Nat. Commun.* **13**, 5792 (2022).

23. Schmitz, M., Querques, I., Oberli, S., Chanez, C. & Jinek, M. Structural basis for the assembly of the type V CRISPR-associated transposon complex. *Cell* https://doi.org/10.1016/j.cell.2022.11.009 (2022).

24. Gary, P. A., Biery, M. C., Bainton, R. J. & Craig, N. L. Multiple DNA processing reactions underlie Tn7 transposition. *J. Mol. Biol.* **257**, 301–316 (1996).

25. Montano, S. P., Pigli, Y. Z. & Rice, P. A. The mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413–417 (2012).

26. Kaczmarska, Z. et al. Structural basis of transposon end recognition explains central features of Tn7 transposition systems. *Mol. Cell* **82**, 2618–2632.e7 (2022).

27. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).

28. Skelding, Z., Queen-Baker, J. & Craig, N. L. Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. *EMBO J.* **22**, 5904–5917 (2003).

29. Jullien, N. & Herman, J. P. LUEGO: a cost and time saving gel shift procedure. *Biotechniques* **51**, 267–269 (2011).

30. Wang, F. et al. General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. *Proc. Natl Acad. Sci. USA* **117**, 24269–24273 (2020).

31. Patel, A., Toso, D., Litvak, A. & Nogales, E. Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from

tilted grids. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.08.434344 (2021).

32. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

33. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

34. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).

35. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

36. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

37. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).

38. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

39. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

40. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

41. Punjani, A. & Fleet, D. J. 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).

42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

43. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

44. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D* **74**, 814–840 (2018).

45. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

46. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

47. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

48. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).

49. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

50. Kaur, S. et al. Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nat. Commun.* **12**, 1240 (2021).

51. Cianfrocco MA, W. M., Youn, C., Wagner, R. & Leschziner, A. E. COSMIC^2: a science gateway for cryo-electron microscopy structure determination. *Pract. Exp. Adv. Res. Comput.* http://doi.acm.org/10.1145/3093338.3093390 (2017).

52. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

Acknowledgements

This work was performed at the National Center for Cryo-EM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy Program (U24 GM129539,) and by grants from the Simons Foundation (SF349247) and NY State Assembly. This work also relied on data collected using an instrument supported by the NIH through award S10OD030470-01. We additionally acknowledge XSEDE for computational resources used for image processing (MCB200090 to E.H.K.). We thank S.-C. Hsieh for sharing the protocol to analyse Illumina sequencing results; J. E. Peters, S. H. Sternberg and A. Guarné along with the A. Guarné, J. E. Peters and E. H. Kellogg laboratories for feedback and stimulating discussion. This research is supported by the NIH: R01GM144566 (E.H.K.) and Pew Biomedical Foundation (E.H.K.). This work relied on data collected using an instrument supported by the NIH through award S10OD030470-01.

Author information

Author notes

1. These authors contributed equally: Jung-Un Park, Amy Wei-Lun Tsai, Alexandrea N. Rizo

Authors and Affiliations

1. Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

   Jung-Un Park, Amy Wei-Lun Tsai, Alexandrea N. Rizo, Vinh H. Truong, Tristan X. Wellner, Richard D. Schargel & Elizabeth H. Kellogg

Contributions

E.H.K. designed and supervised the project. J.-U.P., A.W.-L.T. and T.X.W. carried out protein expression and purification. R.D.S. cloned S15. J.-U.P., A.W.-L.T. and V.H.T prepared sample components and completed DNA-based pulldowns. J.-U.P., A.W.-L.T., V.H.T., A.N.R. and T.X.W. prepared GO grids and optimized the sample for cryo-EM imaging. J.-U.P., A.W.-L.T. and A.N.R. collected, processed, analysed, and refined cryo-EM data. J.-U.P. and A.N.R. built and refined atomic models. J.-U.P. and A.W.-L.T. carried out all in vitro assays (ATP hydrolysis and in vitro transposition assays). J.-U.P., A.W.-L.T., A.N.R. and E.H.K. contributed to figures and manuscript writing. All authors contributed to manuscript writing and editing.

Corresponding author

Correspondence to Elizabeth H. Kellogg.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Designed DNA substrate schematic and negative staining microscopy of target pot and transpososome.

a. DNA substrate design for the transpososome assembly includes two DNA substrates: target pot DNA containing a Cas12k-binding site and the first two TnsB binding sites of left-end (LE), and donor pot DNA with the first two TnsB binding sites of right-end (RE). DNA substrates were designed to form a strand-transfer complex by having 5 base pairs (bp) single-stranded DNA (ssDNA) connecting the target DNA region and the transposon DNA region. The design for target-pot DNA is composed of four single-stranded DNA (ssDNA) of the following: Target-LE_F (light blue), non-target_R (red), desthiobiotinylated LUEGO (green), and LE_R (dark blue). Cas12k-binding region contains PAM and a 10 bp mismatch to facilitate R-loop formation as indicated as a displaced strand. 5′ end labeled desthiobiotin (green circle) on the LUEGO was used to conjugate target-pot DNA on magnetic beads for the pulldown. Second, donor-pot DNA consists of three ssDNA: RE_F (grey), RE_R1 (purple), and RE_R2 (orange). LE_R region and RE_R1 region of each DNA substrate corresponds to the first two TnsB binding sites of LE and RE, respectively. Two DNA substrates have 5 bp of complementary sequences to each other, which are annealed upon transpososome assembly. Locations of PAM, and R-loop are annotated in black. Sequences for all DNA substrates are included in Supplementary Table 1. **b**. Features of the designed DNA substrate. Each left-end and right-end transposon region of the DNA includes 8 bp terminal sequences and the two TnsB binding sites (L1 and L2 for the left end, R1 and R2 for the right end). The beginning of the left-end sequence (red triangle) is 61 bp distant from the PAM as indicated with a black arrow. PAM and the 5 bp of complementary sequences (target site duplication) were represented in red.**c**. Negative stain image of the target pot (i.e. recruitment complex containing Cas12k, S15, TniQ, and TnsC) shows a heterogenous assembly with variable length TnsC filaments (indicated by black arrows). **d**. Negative stain image of the transpososome complex shows the addition of donor pot disassembled TnsC filaments and resulted in a homogeneous sample. Scale bar (white) represents 100 nm. The micrographs shown are examples images from negative-stain screening datasets consisting of 20 and 100 micrographs, respectively.

Extended Data Fig. 2 Cryo-EM imaging and image processing pipeline of the transpososome complex.

a. Representative cryo-EM micrograph from the reconstituted transpososome sample. Scale bar (white, bottom right) represents 100 nm. The micrograph shown is an example image from a full dataset consisting of 17,554 micrographs. **b**. Image processing workflow used to analyze the cryo-EM data. 2D classification in cryoSPARC v3.3.1 on template picked particles (from 14,017 micrographs) resulted in 535,932 particles. *Ab-initio* reconstruction on the initial particle stack resulted in two classes, one with 53% of the particles (pink) and the other with 47% of the particles (gray). The two classes were separated for subsequent classification and refinement steps. Before performing 3D classification in RELION v4, each class underwent homogenous refinement in cryoSPARCv3.3.1[35]. RELION v4 3D classification (without alignments, skip_align)[36,37] was applied to both populations from the *ab-initio* reconstruction resulting in the colored volumes shown (blue, pink, and yellow). On the left, the two classes (pink and yellow) that have the best resolved Cas12k density were combined for downstream refinement to produce the final 3D reconstruction (boxed), which is the major configuration of the transpososome complex with 12 TnsC protomers. Local refinement was performed on three different segments of the map, focusing on: Cas12k (3.1 Å), TnsC (3.2 Å) and TnsB (3.2 Å). On the right, the class that has the best resolved Cas12k and TnsB density (27% of particles, shown in pink) was selected for downstream refinement to produce the final 3D reconstruction (boxed), which is the minor configuration of the transpososome complex with 13 TnsC protomers. Similar local refinement was performed on three different segments of the minor TnsC complex to result in high resolution reconstructions of the target site proteins (Cas12k+TniQ+S15), TnsC, and TnsB. **c**-**d**. Fourier shell correlation (FSC) curve of the major (**c**) and minor (**d**) configuration of the transpososome complex, respectively. Masked (dashed) or unmasked (solid) gold standard half-map FSC (blue) and model-map FSC (red) curves are shown for the refined reconstruction and atomic model. Model-map cutoff (0.5) and gold-standard FSC cutoffs (0.143) are indicated with dashed lines. Estimated resolution based on these cutoffs are indicated. **e**-**f**. Local resolution filtered reconstruction for the major (**e**) and minor (**f**) configuration of the transpososome complex, respectively, are shown with

estimated local resolution indicated using colored surface. Local resolution ranges from 3.0 Å (blue) to 7.0 Å (red). Legend at the bottom indicates local resolution range and values in Angstrom. **g**-**h**. Angular distribution plot for particle projections of the major (**g**) and minor (**h**) configuration of the transpososome complex, respectively. The plot was calculated in cryoSPARC v3.3.1 and shows the number of particles for each viewing angle. Colors indicate counts; red corresponds to high particle counts for that particular viewing angle, blue to low particle counts.

Extended Data Fig. 3 TnsC has dedicated faces for TniQ and TnsB within the transpososome.

A simulated map is colored by different regions of TnsC (opaque surface) to represent the N- and C-terminal face of TnsC. The regions are colored as follows: N-terminal face (residues 19–140 on TnsC7-TnsC12, rose-brown), and C-terminal face (residues 141–275 on TnsC1-TnsC6, light blue). Region not included in either N- or C-terminal face is colored green. TniQ (orange ribbon) and TnsB (purple ribbon) associates with the N-terminal face and C-terminal face respectively.

Extended Data Fig. 4 Interactions between transpososome components near R-loop.

a. DNA (blue ribbon) forms 17 base pair heteroduplex with RNA (gray) in Cas12k (pink). Cas12k, S15 (tan), and TniQ (orange) are shown in surface representation. Close-up view on the right shows the model docked into the cryo-EM density of the R-loop. The PAM distal end of the R-loop is indicated with an asterisk (*). **b**. Atomic model of the ShCAST transpososome, focusing on Cas12k. The TnsC protomer closest to Cas12k (TnsC10) is labeled. TnsC protomers are numbered as previously defined (see Main Text, Fig. 2). ShCAST protein and nucleic acid components are labeled and colored according to previously defined colors (see Main Text, Fig. 1). Black box indicates the TnsC finger loop that is close to Cas12k shown as inset in panel B. **c**. The closet distance between TnsC protomer TnsC10 and Cas12k is shown with dashed line and labeled (in Å). **d**. S15 (beige) is positioned between the REC2 domain of Cas12k (pink(and the sgRNA-DNA heteroduplex (blue/grey). The rooftop loop of sgRNA (grey)

is stabilized by S15 (beige) and TniQ (orange). Rotation with respect to panel a is indicated in top left corner.

Extended Data Fig. 5 TnsB^{Hook} density occupies select TnsC protomers in the TnsC hexamer closest to TnsB.

a. Local resolution filtered cryo-EM reconstruction (same as that shown in Fig. 1) colored by assignment reveals that TnsB^{Hook} (light and dark purple) occupies binding sites on TnsC (green). **b**. Rotation of the cryo-EM reconstruction by 30° shows the adjacent TnsB^{Hook} binding pocket on TnsC (white dashed lines) is empty.

Extended Data Fig. 6 High throughput mapping of the *in vitro* transposition events reveals that the identified interaction between TnsC and TnsB IIβ domain is crucial for target-site selection.

Insertion positions were determined by Illumina sequencing of the plasmids extracted from colonies from *in vitro* transpositions under the following conditions (same with Fig. 3f, see Methods): **a**. TnsB wild-type (WT) with S15, **b**. TnsB WT without S15, **c**. TnsB R432A with S15, **d**. TnsB R432A without S15, **e**. TnsB Y439A with S15, and **f**. TnsB Y439A without S15. Determined insertion positions were plotted as a histogram indicating the percentage of the reads at the 10 base-pair (bp) windows within the target plasmid. Position numbers (x-axis) correspond to the number of base pairs between the PAM and the beginning of the transposon-end sequence after the transposition. Red and blue bars represent the transposition products with the left end-right end (L-R, correct) or the right end-left end (R-L, wrong) orientation respectively. For the conditions with high on-target percentage (> 60%, panels a, b, c, and e), insets are presented for the positions around the PSP1 protospacer (from 0 bp to 70 bp), which is indicated with black brackets on the x-axis. Grey bar in the inset indicates the 17 bp PSP1 protospacer. For panel e, a red bar on the y-axis represents the region for zoom-in on the lower panel to visualize signals from the off-target transposition events. Two origins of replications within the target

plasmids (f1 ori and ori) are annotated as black bars on the x-axis, which explains the reason for the cold spots for the transpositions.

Extended Data Fig. 7 TnsC in ShCAST transpososome does not match helical parameters of the bound DNA.

The repeat length of TnsC turn (~6 protomers per turn) is approximately 40 Å, while the repeat length of the TnsC-bound DNA (~11 base pairs per turn) is approximately 36 Å. DNA model is represented in a solid ribbon. Protein components and sgRNA in the transpososome are represented as transparent surfaces. Color scheme is identical to the established colors in Fig. 1.

Extended Data Fig. 8 TnsC-DNA interaction proximal to TnsB is identical in the major and minor configurations.

a. Low pass filtered (10 Å) cryo-EM reconstructions of both major (transparent grey) and minor (solid, colored) configurations are aligned with respect to TnsB. The dashed box indicates the TnsC-DNA interactions at the TnsB proximal region shown as inset in panel B. **b**. Three TnsB-proximal TnsC protomers (From TnsC1 to TnsC3) in both major and minor configurations interact with target strand DNA (tsDNA) in an identical manner through residue R182. Hydrogen bonding interactions (distance cut off <4 Å) between protein residues and the sugar-phosphate backbone of DNA are represented with dashed lines. Base pairs that are interacting with TnsC are represented as filled nucleotides and stick phosphate-backbone.

Extended Data Table 1 Cryo-EM Map and Model Validation
Extended Data Table 2 The distance between TnsC residues and the closest atom from the phosphate backbone of DNA

Supplementary information

Supplementary Information

Supplementary Figs. 1–11; Table 1, which includes all the oligonucleotides used in this study; and references.

[Reporting Summary]

[Peer Review File]

[Supplementary Video 1]

3D variability analysis visualizes a flexible association of Cas12k with respect to the rest of the transpososome. Components in the transpososome were coloured following the convention established in Fig. 1. See Methods for details.

[Supplementary Video 2]

Conformational change between TnsC filaments in the major configuration of transpososome and helical TnsC. The movie shows the morphing between twelve protomers of TnsC from the major configuration of transpososome and helical TnsC structure (PDB: 7M99).

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Park, JU., Tsai, A.WL., Rizo, A.N. *et al.* Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature* **613**, 775–782 (2023). https://doi.org/10.1038/s41586-022-05573-5

- Received: 30 September 2022
- Accepted: 17 November 2022
- Published: 28 November 2022
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05573-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **First full views of a CRISPR-guided system for gene insertion**
  - Orsolya Barabas
  - Phoebe A. Rice

  Nature (2023)

First full views of a CRISPR-guided system for gene insertion

- Orsolya Barabas
- Phoebe A. Rice

Nature News & Views 11 Jan 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05573-5

| Section menu | Main menu |

Amendments & Corrections

- **Author Correction: Inner ear biomechanics reveals a Late Triassic origin for mammalian endothermy** [09 Jan 2023]
  -
- **Author Correction: Androgen receptor blockade promotes response to BRAF/MEK-targeted therapy** [10 Jan 2023]
  -
- **Author Correction: Steady Floquet–Andreev states in graphene Josephson junctions** [10 Jan 2023]
  -
- **Publisher Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry** [09 Jan 2023]
  -
- **Publisher Correction: A very luminous jet from the disruption of a star by a massive black hole** [09 Jan 2023]
  -

- Author Correction
- Published: 09 January 2023

Author Correction: Inner ear biomechanics reveals a Late Triassic origin for mammalian endothermy

- Ricardo Araújo ,
- Romain David ,
- Julien Benoit,
- Jacqueline K. Lungmus,
- Alexander Stoessel,
- Paul M. Barrett,
- Jessica A. Maisano,
- Eric Ekdale,
- Maëva Orliac,
- Zhe-Xi Luo,
- Agustín G. Martinelli,
- Eva A. Hoffman,
- Christian A. Sidor,
- Rui M. S. Martins,
- Fred Spoor &
- …
- Kenneth D. Angielczyk

Show authors

Nature **volume 613**, page E2 (2023)

- 437 Accesses

- 2 Altmetric
- Metrics details

The Original Article was published on 20 July 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-04963-z
Published online 20 July 2022

In the version of this article initially published, the Acknowledgements did not include thanks to the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities and Vincent Fernandez and Paul Tafforeau for assistance in using beamline BM05 and ID17. The omissions have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Ricardo Araújo, Romain David

Authors and Affiliations

1. Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

   Ricardo Araújo & Rui M. S. Martins

2. Institut des Sciences de L'Évolution de Montpellier, Université de Montpellier, Montpellier, France

   Ricardo Araújo & Maëva Orliac

3. Natural History Museum, London, UK

Romain David, Paul M. Barrett & Fred Spoor

4. Max Plank Institute for Evolutionary Anthropology, Leipzig, Germany

Romain David, Alexander Stoessel & Fred Spoor

5. Evolutionary Studies Institute, University of Witwatersrand, Johannesburg, South Africa

Julien Benoit

6. Department of Paleobiology, National Museum of Natural History, Washington DC, USA

Jacqueline K. Lungmus

7. Institute of Zoology and Evolutionary Research, Friedrich Schiller University Jena, Jena, Germany

Alexander Stoessel

8. Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA

Jessica A. Maisano

9. Department of Biology, San Diego State University, San Diego, CA, USA

Eric Ekdale

10. Department of Paleontology, San Diego Natural History Museum, San Diego, CA, USA

Eric Ekdale

11. Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA

Zhe-Xi Luo

12. Museo Argentino de Ciencias Naturales ‘Bernardino Rivadavia’, Buenos Aires, Argentina

    Agustín G. Martinelli

13. Division of Paleontology, American Museum of Natural History, New York, NY, USA

    Eva A. Hoffman

14. Burke Museum and Department of Biology, University of Washington, Seattle, WA, USA

    Christian A. Sidor

15. Department of Anthropology, University College London, London, UK

    Fred Spoor

16. Neguanee Integrative Research Center, Field Museum of Natural History, Chicago, IL, USA

    Kenneth D. Angielczyk

Corresponding authors

Correspondence to Ricardo Araújo, Romain David or Kenneth D. Angielczyk.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Araújo, R., David, R., Benoit, J. *et al.* Author Correction: Inner ear biomechanics reveals a Late Triassic origin for mammalian endothermy. *Nature* **613**, E2 (2023). https://doi.org/10.1038/s41586-023-05700-w

- Published: 09 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-023-05700-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05700-w

- Author Correction
- Published: 10 January 2023

Author Correction: Androgen receptor blockade promotes response to BRAF/MEK-targeted therapy

- Christopher P. Vellano,
- Michael G. White,
- Miles C. Andrews,
- Manoj Chelvanambi,
- Russell G. Witt,
- Joseph R. Daniele,
- Mark Titus,
- Jennifer L. McQuade,
- Fabio Conforti,
- Elizabeth M. Burton,
- Matthew J. Lastrapes,
- Gabriel Ologun,
- Alexandria P. Cogdill,
- Golnaz Morad,
- Peter Prieto,
- Alexander J. Lazar,
- Yanshuo Chu,
- Guangchun Han,
- M. A. Wadud Khan,
- Beth Helmink,
- Michael A. Davies,
- Rodabe N. Amaria,

- Jeffrey J. Kovacs,
- Scott E. Woodman,
- Sapna Patel,
- Patrick Hwu,
- Michael Peoples,
- Jeffrey E. Lee,
- Zachary A. Cooper,
- Haifeng Zhu,
- Guang Gao,
- Hiya Banerjee,
- Mike Lau,
- Jeffrey E. Gershenwald,
- Anthony Lucci,
- Emily Z. Keung,
- Merrick I. Ross,
- Laura Pala,
- Eleonora Pagan,
- Rossana Lazcano Segura,
- Qian Liu,
- Mikayla S. Borthwick,
- Eric Lau,
- Melinda S. Yates,
- Shannon N. Westin,
- Khalida Wani,
- Michael T. Tetzlaff,
- Lauren E. Haydu,
- Mikhila Mahendra,
- XiaoYan Ma,
- Christopher Logothetis,
- Zachary Kulstad,
- Sarah Johnson,
- Courtney W. Hudgens,
- Ningping Feng,
- Lorenzo Federico,
- Georgina V. Long,
- P. Andrew Futreal,
- Swathi Arur,

- Hussein A. Tawbi,
- Amy E. Moran,
- Linghua Wang,
- Timothy P. Heffernan ,
- Joseph R. Marszalek &
- …
- Jennifer A. Wargo

Show authors

Nature **volume 613**, page E3 (2023)

- 500 Accesses

- 1 Altmetric

- Metrics details

The Original Article was published on 15 June 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-04833-8
Published online 15 June 2022

In the version of this article initially published, author affiliation 5, now reading "Division of Melanoma, Sarcoma, and Rare Tumors, European Institute of Oncology, IRCCS, Milan, Italy," incorrectly included "Medical Oncology for" in the name of the division, and the IRCCS was not included. The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. Miles C. Andrews

Present address: Department of Medicine, Central Clinical School, Monash University, Melbourne, Victoria, Australia

2. Gabriel Ologun

   Present address: Department of Surgery, Guthrie Courtland Medical Center, Courtland, NY, USA

3. Alexandria P. Cogdill

   Present address: Immunai, New York, NY, USA

4. Peter Prieto

   Present address: Department of Surgery, University of Rochester, Rochester, NY, USA

5. Beth Helmink

   Present address: Department of Surgery, Washington University in St Louis, St Louis, MO, USA

6. Patrick Hwu

   Present address: Moffitt Cancer Center, Tampa, FL, USA

7. Zachary A. Cooper

   Present address: AstraZeneca, Gaithersburg, MD, USA

8. Michael T. Tetzlaff

   Present address: Department of Pathology, University of California, San Francisco, CA, USA

9. These authors contributed equally: Christopher P. Vellano, Michael G. White, Miles C. Andrews, Manoj Chelvanambi, Russell G. Witt

Authors and Affiliations

1. TRACTION Platform, Therapeutics Discovery Division, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Christopher P. Vellano, Joseph R. Daniele, Jeffrey J. Kovacs, Michael Peoples, Guang Gao, Mikhila Mahendra, XiaoYan Ma, Ningping Feng, Lorenzo Federico, Timothy P. Heffernan & Joseph R. Marszalek

2. Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Michael G. White, Miles C. Andrews, Manoj Chelvanambi, Russell G. Witt, Elizabeth M. Burton, Matthew J. Lastrapes, Gabriel Ologun, Alexandria P. Cogdill, Golnaz Morad, Peter Prieto, M. A. Wadud Khan, Beth Helmink, Jeffrey E. Lee, Zachary A. Cooper, Jeffrey E. Gershenwald, Anthony Lucci, Emily Z. Keung, Merrick I. Ross, Lauren E. Haydu, Zachary Kulstad, Sarah Johnson & Jennifer A. Wargo

3. Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Mark Titus & Christopher Logothetis

4. Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Jennifer L. McQuade, Michael A. Davies, Rodabe N. Amaria, Sapna Patel, Patrick Hwu & Hussein A. Tawbi

5. Division of Melanoma, Sarcoma, and Rare Tumors, European Institute of Oncology, IRCCS, Milan, Italy

   Fabio Conforti & Laura Pala

6. Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

   Alexander J. Lazar & Michael T. Tetzlaff

7. Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Alexander J. Lazar, Yanshuo Chu, Guangchun Han, Scott E. Woodman, Haifeng Zhu, P. Andrew Futreal, Linghua Wang & Jennifer A. Wargo

8. Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Alexander J. Lazar, Rossana Lazcano Segura, Khalida Wani & Courtney W. Hudgens

9. Clinical Development and Analytics, Novartis Pharmaceuticals, East Hanover, NJ, USA

Hiya Banerjee

10. Novartis Pharma, Basel, Switzerland

Mike Lau

11. Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy

Eleonora Pagan

12. Department of Tumor Biology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

Qian Liu & Eric Lau

13. Department of Gynecologic Oncology and Reproductive Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Mikayla S. Borthwick, Melinda S. Yates & Shannon N. Westin

14. Melanoma Institute Australia, The University of Sydney, and Royal North Shore and Mater Hospitals, Sydney, New South Wales, Australia

Georgina V. Long

15. Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Swathi Arur

16. Cell, Development & Cancer Biology, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA

Amy E. Moran

Corresponding authors

Correspondence to Timothy P. Heffernan, Joseph R. Marszalek or Jennifer A. Wargo.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Vellano, C.P., White, M.G., Andrews, M.C. *et al.* Author Correction: Androgen receptor blockade promotes response to BRAF/MEK-targeted therapy. *Nature* **613**, E3 (2023). https://doi.org/10.1038/s41586-022-05632-x

- Published: 10 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05632-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05632-x

- Author Correction
- Published: 10 January 2023

Author Correction: Steady Floquet–Andreev states in graphene Josephson junctions

- Sein Park,
- Wonjun Lee,
- Seong Jang,
- Yong-Bin Choi,
- Jinho Park,
- Woochan Jung,
- Kenji Watanabe,
- Takashi Taniguchi,
- Gil Young Cho &
- …
- Gil-Ho Lee

Show authors

Nature **volume 613**, page E4 (2023)

- 344 Accesses
- 2 Altmetric
- Metrics details

The Original Article was published on 16 March 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-021-04364-8
Published online 16 March 2022

In the version of this article initially published, the Acknowledgements omitted thanks to the National Research Foundation of Korea for supporting grant no. 2021R1A6A1A10042944. The error has been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Sein Park, Wonjun Lee

Authors and Affiliations

1. Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea

   Sein Park, Wonjun Lee, Seong Jang, Yong-Bin Choi, Jinho Park, Woochan Jung, Gil Young Cho & Gil-Ho Lee

2. Center for Artificial Low Dimensional Electronic Systems, Institute for Basic Science (IBS), Pohang, Republic of Korea

   Wonjun Lee & Gil Young Cho

3. Research Center for Functional Materials, National Institute for Materials Science, Tsukuba, Japan

   Kenji Watanabe

4. International Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba, Japan

   Takashi Taniguchi

5. Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea

Gil Young Cho & Gil-Ho Lee

Corresponding authors

Correspondence to Gil Young Cho or Gil-Ho Lee.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Park, S., Lee, W., Jang, S. *et al.* Author Correction: Steady Floquet–Andreev states in graphene Josephson junctions. *Nature* **613**, E4 (2023). https://doi.org/10.1038/s41586-023-05709-1

- Published: 10 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-023-05709-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05709-1

| Section menu | Main menu |

- Publisher Correction
- Published: 09 January 2023

Publisher Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry

- John Kee,
- Samuel Thudium,
- David M. Renner,
- Karl Glastad,
- Katherine Palozola,
- Zhen Zhang,
- Yize Li,
- Yemin Lan,
- Joseph Cesare,
- Andrey Poleshko,
- Anna A. Kiseleva,
- Rachel Truitt,
- Fabian L. Cardenas-Diaz,
- Xianwen Zhang,
- Xuping Xie,
- Darrell N. Kotton,
- Konstantinos D. Alysandratos,
- Jonathan A. Epstein,
- Pei-Yong Shi,
- Wenli Yang,
- Edward Morrisey,
- Benjamin A. Garcia,
- Shelley L. Berger,
- Susan R. Weiss &
- …

- Erica Korb

Show authors

Nature **volume 613**, page E5 (2023)

- 561 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 05 October 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05282-z
Published online 5 October 2022

In the version of this article initially published, the top-left label of Fig. 2d, now reading “H3K27me3,” appeared incorrectly as “H3K9me3,” and the top *x*-axis label of the left panel in Fig. 4a, now reading “SARS-CoV-2^{WT},” appeared incorrectly as “SARS-CoV-2$^{\Delta ARKSAP}$.” The figures have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Samuel Thudium, David M. Renner, Karl Glastad

Authors and Affiliations

1. Department of Genetics at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   John Kee, Samuel Thudium, Katherine Palozola & Erica Korb

2. Epigenetics Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   John Kee, Samuel Thudium, Karl Glastad, Katherine Palozola, Zhen Zhang, Yemin Lan, Joseph Cesare, Benjamin A. Garcia, Shelley L. Berger & Erica Korb

3. Department of Microbiology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   David M. Renner, Yize Li & Susan R. Weiss

4. Penn Center for Research on Coronaviruses and Other Emerging Pathogens at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   David M. Renner, Yize Li & Susan R. Weiss

5. Department of Cell and Developmental Biology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Karl Glastad, Zhen Zhang, Andrey Poleshko, Anna A. Kiseleva, Jonathan A. Epstein & Shelley L. Berger

6. Department of Biochemistry and Biophysics at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Joseph Cesare & Benjamin A. Garcia

7. Penn Cardiovascular Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Anna A. Kiseleva & Jonathan A. Epstein

8. Institute for Regenerative Medicine at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Anna A. Kiseleva & Jonathan A. Epstein

9. Department of Medicine at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

   Rachel Truitt, Fabian L. Cardenas-Diaz, Jonathan A. Epstein, Wenli Yang & Edward Morrisey

10. Penn-CHOP Lung Biology Institute at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

    Fabian L. Cardenas-Diaz & Edward Morrisey

11. Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

    Xianwen Zhang, Xuping Xie & Pei-Yong Shi

12. Center for Regenerative Medicine, Boston University and Boston Medical Center, Boston, MA, USA

    Darrell N. Kotton & Konstantinos D. Alysandratos

13. The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA, USA

    Darrell N. Kotton & Konstantinos D. Alysandratos

14. Department of Biology at the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

    Shelley L. Berger

Corresponding author

Correspondence to Erica Korb.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Kee, J., Thudium, S., Renner, D.M. *et al.* Publisher Correction: SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry. *Nature* **613**, E5 (2023). https://doi.org/10.1038/s41586-022-05631-y

- Published: 09 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05631-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05631-y

- Publisher Correction
- Published: 09 January 2023

Publisher Correction: A very luminous jet from the disruption of a star by a massive black hole

- Igor Andreoni ,
- Michael W. Coughlin ,
- Daniel A. Perley,
- Yuhan Yao,
- Wenbin Lu,
- S. Bradley Cenko,
- Harsh Kumar,
- Shreya Anand,
- Anna Y. Q. Ho,
- Mansi M. Kasliwal,
- Antonio de Ugarte Postigo,
- Ana Sagués-Carracedo,
- Steve Schulze,
- D. Alexander Kann,
- S. R. Kulkarni,
- Jesper Sollerman,
- Nial Tanvir,
- Armin Rest,
- Luca Izzo,
- Jean J. Somalwar,
- David L. Kaplan,
- Tomás Ahumada,
- G. C. Anupama,
- Katie Auchettl,
- Sudhanshu Barway,

- Eric C. Bellm,
- Varun Bhalerao,
- Joshua S. Bloom,
- Michael Bremer,
- Mattia Bulla,
- Eric Burns,
- Sergio Campana,
- Poonam Chandra,
- Panos Charalampopoulos,
- Jeff Cooke,
- Valerio D'Elia,
- Kaustav Kashyap Das,
- Dougal Dobie,
- José Feliciano Agüí Fernández,
- James Freeburn,
- Cristoffer Fremling,
- Suvi Gezari,
- Simon Goode,
- Matthew J. Graham,
- Erica Hammerstein,
- Viraj R. Karambelkar,
- Charles D. Kilpatrick,
- Erik C. Kool,
- Melanie Krips,
- Russ R. Laher,
- Giorgos Leloudas,
- Andrew Levan,
- Michael J. Lundquist,
- Ashish A. Mahabal,
- Michael S. Medford,
- M. Coleman Miller,
- Anais Möller,
- Kunal P. Mooley,
- A. J. Nayana,
- Guy Nir,
- Peter T. H. Pang,
- Emmy Paraskeva,

- Richard A. Perley,
- Glen Petitpas,
- Miika Pursiainen,
- Vikram Ravi,
- Ryan Ridden-Harper,
- Reed Riddle,
- Mickael Rigault,
- Antonio C. Rodriguez,
- Ben Rusholme,
- Yashvi Sharma,
- I. A. Smith,
- Robert D. Stein,
- Christina Thöne,
- Aaron Tohuvavohu,
- Frank Valdes,
- Jan van Roestel,
- Susanna D. Vergani,
- Qinan Wang &
- …
- Jielai Zhang

Show authors

Nature **volume 613**, page E6 (2023)

- 366 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 30 November 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05465-8
Published online 30 November 2022

In the version of this article initially published, there was in an error in the third-to-last sentence of the abstract, now reading, in part, “we calculate a rate of \({0.02}_{\mbox{--}0.01}^{+0.04}\) Gpc^{-3} yr^{-1}”, where Gpc was spelled out as gigapascals, not gigaparsecs. Also, the scale label (2″) was missing in the lower-left corner of Fig. 1b. The errors have been corrected in the HTML and PDF versions of the article.

Author information

Author notes

1. These authors contributed equally: Igor Andreoni and Michael W. Coughlin

Authors and Affiliations

1. Joint Space-Science Institute, University of Maryland, College Park, MD, USA

   Igor Andreoni, S. Bradley Cenko & M. Coleman Miller

2. Department of Astronomy, University of Maryland, College Park, MD, USA

   Igor Andreoni, Tomás Ahumada, Erica Hammerstein & M. Coleman Miller

3. Astrophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA

   Igor Andreoni & S. Bradley Cenko

4. School of Physics and Astronomy, University of Minnesota, Minneapolis, MN, USA

   Michael W. Coughlin

5. Astrophysics Research Institute, Liverpool John Moores University, Liverpool, UK

Daniel A. Perley

6. Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA, USA

Yuhan Yao, Shreya Anand, Mansi M. Kasliwal, S. R. Kulkarni, Jean J. Somalwar, Kaustav Kashyap Das, Cristoffer Fremling, Matthew J. Graham, Viraj R. Karambelkar, Ashish A. Mahabal, Kunal P. Mooley, Vikram Ravi, Antonio C. Rodriguez, Yashvi Sharma, Robert D. Stein & Jan van Roestel

7. Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

Wenbin Lu

8. Indian Institute of Technology Bombay, Mumbai, India

Harsh Kumar & Varun Bhalerao

9. Department of Astronomy, University of California, Berkeley, Berkeley, CA, USA

Anna Y. Q. Ho, Joshua S. Bloom, Michael S. Medford & Guy Nir

10. Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Anna Y. Q. Ho, Joshua S. Bloom & Michael S. Medford

11. Miller Institute for Basic Research in Science, Berkeley, CA, USA

Anna Y. Q. Ho

12. Artemis, Université Côte d’Azur, Observatoire de la Côte d’Azur, CNRS, Nice, France

Antonio de Ugarte Postigo

13. The Oskar Klein Centre, Department of Physics, Stockholm University, AlbaNova, Stockholm, Sweden

Ana Sagués-Carracedo & Steve Schulze

14. Instituto de Astrofísica de Andalucía, Granada, Spain

D. Alexander Kann & José Feliciano Agüí Fernández

15. The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, Stockholm, Sweden

Jesper Sollerman, Mattia Bulla & Erik C. Kool

16. Department of Physics and Astronomy, University of Leicester, Leicester, UK

Nial Tanvir

17. Space Telescope Science Institute, Baltimore, MD, USA

Armin Rest, Suvi Gezari & Qinan Wang

18. Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD, USA

Armin Rest

19. DARK, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Luca Izzo

20. Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

David L. Kaplan

21. Indian Institute of Astrophysics, Bangalore, India

G. C. Anupama & Sudhanshu Barway

22. School of Physics, University of Melbourne, Parkville, Victoria, Australia

Katie Auchettl

23. ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Sydney, Australia

Katie Auchettl

24. Department of Astronomy and Astrophysics, University of California, Santa Cruz, Santa Cruz, CA, USA

Katie Auchettl

25. DIRAC Institute, Department of Astronomy, University of Washington, Seattle, WA, USA

Eric C. Bellm

26. Institut de Radioastronomie Millimétrique (IRAM), Saint Martin d'Hères, France

Michael Bremer & Melanie Krips

27. Department of Physics & Astronomy, Louisiana State University, Baton Rouge, LA, USA

Eric Burns

28. INAF-Osservatorio Astronomico di Brera, Merate, Italy

Sergio Campana

29. National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Pune, India

Poonam Chandra

30. DTU Space, National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark

Panos Charalampopoulos, Giorgos Leloudas & Miika Pursiainen

31. Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), Swinburne University of Technology, Hawthorn, Victoria, Australia

Jeff Cooke, Dougal Dobie, James Freeburn, Simon Goode, Anais Möller & Jielai Zhang

32. Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria, Australia

Jeff Cooke, Dougal Dobie, James Freeburn, Simon Goode, Anais Möller & Jielai Zhang

33. Space Science Data Center – Agenzia Spaziale Italiana, Rome, Italy

Valerio D'Elia

34. Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL, USA

Charles D. Kilpatrick

35. IPAC, California Institute of Technology, Pasadena, CA, USA

Russ R. Laher & Ben Rusholme

36. Department of Astrophysics, Radboud University, Nijmegen, The Netherlands

Andrew Levan

37. W. M. Keck Observatory, Kamuela, HI, USA

Michael J. Lundquist

38. Center for Data-Driven Discovery, California Institute of Technology, Pasadena, CA, USA

Ashish A. Mahabal

39. Indian Institute of Astrophysics, Bengaluru, India

A. J. Nayana

40. Nikhef, Amsterdam, The Netherlands

Peter T. H. Pang

41. Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Utrecht, The Netherlands

Peter T. H. Pang

42. IAASARS, National Observatory of Athens, Penteli, Greece

Emmy Paraskeva

43. Department of Astrophysics, Astronomy and Mechanics, Faculty of Physics, National and Kapodistrian University of Athens, Athens, Greece

Emmy Paraskeva

44. Nordic Optical Telescope, Breña Baja, Spain

Emmy Paraskeva

45. Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark

Emmy Paraskeva

46. National Radio Astronomy Observatory, Socorro, NM, USA

Richard A. Perley

47. Center for Astrophysics – Harvard & Smithsonian, Cambridge, MA, USA

Glen Petitpas

48. School of Physical and Chemical Sciences | Te Kura Matū, University of Canterbury, Christchurch, New Zealand

Ryan Ridden-Harper

49. Caltech Optical Observatories, California Institute of Technology, Pasadena, CA, USA

Reed Riddle

50. Université Lyon, Université Claude Bernard Lyon 1, CNRS, IP2I Lyon / IN2P3, IMR 5822, Villeurbanne, France

Mickael Rigault

51. Institute for Astronomy, University of Hawaiʻi, Pukalani, HI, USA

I. A. Smith

52. Astronomical Institute of the Czech Academy of Sciences (ASU-CAS), Ondřejov, Czech Republic

Christina Thöne

53. Department of Astronomy and Astrophysics, University of Toronto, Toronto, Ontario, Canada

Aaron Tohuvavohu

54. National Optical-Infrared Astronomy Research Laboratory, Tucson, AZ, USA

Frank Valdes

55. GEPI, Observatoire de Paris, PSL University, CNRS, Meudon, France

Susanna D. Vergani

56. Institut d'Astrophysique de Paris, CNRS-UPMC, UMR7095, Paris, France

Susanna D. Vergani

Corresponding authors

Correspondence to Igor Andreoni or Michael W. Coughlin.

Additional information

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Andreoni, I., Coughlin, M.W., Perley, D.A. *et al.* Publisher Correction: A very luminous jet from the disruption of a star by a massive black hole. *Nature* **613**, E6 (2023). https://doi.org/10.1038/s41586-023-05699-0

- Published: 09 January 2023
- Issue Date: 26 January 2023
- DOI: https://doi.org/10.1038/s41586-023-05699-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05699-0